

العربيَّة والذَّكاء الاصطناعيّ

مباحث لغوية 09

تحرير
د. المُعتزّ بالله السَّعيد

الباحثون

د. محمد عطية
د. أحمد راغب
د. المُعتزّ بالله السَّعيد
د. نعيم عبد الغني

مباحث لغوية ٥٩

العربيَّة والذَّكاء الاصطناعيّ

تحرير

د. المُعتزّ بالله السَّعيد

الباحثون

د.محمد عطية
د. أحمد راغب
د. المُعتزّ بالله السَّعيد
د. نعيم عبد الغني

١٤٤١هـ - ٢٠١٩م





العربيَّة والذَّكاء الاصطناعيّ

الطبعة الأولى
١٤٤١ هـ - ٢٠١٩ م

المملكة العربية السعودية – الرياض
ص.ب ١٢٥٠٠ الرياض ١١٤٧٣
هاتف:٠٠٩٦٦١١٢٥٨٧٢٦٨ – ٠٠٩٦٦١١٢٥٨١٠٨٢
البريد الإليكتروني: nashr@kaica.org.sa

مركز الملك عبدالله بن عبدالعزيز الدولي لخدمة اللغة العربية، ١٤٤١هـ.
فهرسة مكتبة الملك فهد الوطنية أثناء النشر
السعيد، المعتز بالله
العربية والذكاء الاصطناعي./ المعتز بالله السعيد .-الرياض، ١٤٤١هـ
..ص؛ .. سم
ردمك: ٤ -٦٢ -٨٢٢١ -٦٠٣ -٩٧٨
١ – اللغة العربية – معالجة البيانات ٢- الذكاء الاصطناعي أ. العنوان
ديوي ٤١٠,٢٨٥ ١٤٤١/٤٦٢
رقم الإيداع: ١٤٤١/٤٦٢
ردمك: ٤ -٦٢ -٨٢٢١ -٦٠٣ -٩٧٨

التصميم والإخراج
دار وجوه للنشر والتوزيع
Wojooh Publishing & Distribution House
www.wojoooh.com
المملكة العربية السعودية – الرياض
الهاتف:4562410 الفاكس:4561675
للتواصل والنشر:
info@wojoooh.com



هذا المشروع

مشروع تأليف سلسلة كتب في مجال (حوسبة العربية) يهدف إلى بناء تراكم معرفي في مجال حيوي مهم، هو مجال (حوسبة العربية) . ويعد هذا الكتاب واحدا من سلسلة كتب صدرت في المركز.

يقع هذا المشروع ضمن سلسلة (مباحث لغوية) التي يشرف المركز على اختيار عنواناتها، وتكليف المحررين والمؤلفين، ومتابعة التأليف حتى إصدار الكتب. وهي سلسلة يجتهد المركز أن تكون سداداً لحاجات بحثية وعلمية تحتاج إلى تنبيه الباحثين عليها، أو تكثيف البحث فيها.

ويعدّ هذا الكتاب واحداً من كتب ثلاثة مترابطة في مشروع علمي واحد متخصص في (الذكاء الاصطناعي) :

- العربية والذكاء الاصطناعي.
- تطبيقات الذكاء الاصطناعي في خدمة اللغة العربية.
- خوارزميات الذكاء الاصطناعي في تحليل النص العربي.

مدير مشروع (العربية والذكاء الاصطناعي)

د.عبدالله بن يحيى الفيفي

فهرس الكتاب

كلمة المركز

يعمل المركز في مجال البحث العلمي ونشر الكتب مستهدفاً التركيز على المجالات البحثية التي ما زالت بحاجة إلى تسليط الضوء عليها، وتكثيف البحث فيها، ولفت أنظار الباحثين والجهات الأكاديمية إلى أهمية استثمارها بمختلف وجوه الاستثمار، وذلك مثل مجال (التخطيط اللغوي) و (العربية في العالم) و(الأدلة والمعلومات) و (تعليم العربية لأبنائها أو لغير الناطقين بها) إلى غير ذلك من المجالات، وإن من أهم مجالات البحث المستقبلية في اللغة العربية مجال (العربية والحوسبة ، والذكاء الاصطناعي) حيث إن حياة اللغات ومستقبلها مرهونة بمدى تجاوبها مع التطورات التقنية والعالم الافتراضي، وكثافة المحتوى الالكتروني المكتوب، وهو ما يشكّل تحديا حقيقيا أمام اللغات غير المنتجة للمعرفة أو للتقنية.

وقد عمل المركز على تسليط الضوء على هذا المجال التخصصي؛ مستعينا بالكفاءات القادرة من المهتمين بالتخصص البيني (بين اللغة والحاسوب) مقدّرا جهودهم، وهادفاً إلى نشرها، وتعميم مبادئها، راغباً أن يكون هذا المسار العلمي مقررا في الجامعات في كلية العربية والحاسوب، ومجالا بحثيا يقصده الباحثون الأكاديميون، والجهات البحثية العربية.

وقد أصدر المركز سابقا ستة عشر كتاباً مختصا في (حوسبة العربية) وفي الإفادة من (المدونات اللغوية) في الأبحاث العربية، ويحتفل بإصدار سبعة كتب جديدة مختصة في (حوسبة العربية والذكاء الاصطناعي)، ويقدمها للقارئ العربي، وللجهات الأكاديمية؛ للإفادة منها في مناهج التعليم والبناء عليه، وهذه الكتب السبعة هي: (العربيَّة والذَّكاء الاصطناعيّ، تطبيقات الذكاء الاصطناعي في خدمة اللغة العربية، خوارزميات الذكاء الاصطناعي في تحليل النص العربي، مقدمة في حوسبة اللغة العربية، الموارد اللغوية الحاسوبية، المعالجة الآلية للنصوص العربية، تطبيقات أساسية في المعالجة الآلية للغة العربية).

ويشكر المركز السادة مؤلفي الكتب، ومحرريها، لما تفضلوا به من عمل علمي رصين، وأدعو الباحثين والمؤلفين إلى التواصل مع المركز لاستكمال المسيرة، وتفتيق فضاءات المعرفة.

وفق الله الجهود وسدد الرؤى.

الأمين العام

أ. د. محمود إسماعيل صالح

مُقدِّمة

يُمثِّلُ «الذَّكاءُ الاصطناعيّ Artificial Intelligence» مَيدانًا بحثيًّا رَحبًا ذا أبعادٍ اقتصاديَّةٍ فائقة، تظهرُ آثارُها على أصعدةٍ مُختلفةٍ وفي ميادينَ شتَّى، تشملُ جوانبَ البحثِ والصِّناعةِ والتَّدريس. ولا يزالُ الأملُ قائمًا لدى المعنيِّينَ بمُستقبل الآلة في الوُصُول بالذَّكاء الاصطناعيّ إلى مُستوياتٍ مُتقدِّمةٍ، يُمكنُ معها توفيرُ كثيرٍ من الجهود والطَّاقات البشريَّة، من خلال تنمية قُدرات الآلة على مُحاكاة ذكاء الإنسان. ولعلَّ الطَّفرةَ الهائلةَ الَّتي نشهدُها اليومَ تُعجِّلُ بتحقيق الآمال والطُّموحات.

لقد مرَّ الذَّكاءُ الاصطناعيُّ - بمفهومه الحديث - بثلاثة أطوارٍ رئيسة، تتمثَّلُ في: طور النَّشأة، وطَور التَّجريب، وطَور النَّهضة الَّتي نَلمَسُها الآن. وكانَت لكلِّ طَورٍ من هذه الأطوار سِماتٌ ورُؤًى وأفكار، تتراكمُ في أُطُرِها العامَّة سعيًا إلى توجيه الآلة إلى فهمٍ أعمقَ لقُدرات العقل البشريّ؛ أو بعبارةٍ أخرى، سعيًا إلى مُحاكاة قُدرات الدِّماغ البشريَّة ومهارات التَّفكير العُليا لدى الإنسان، كالقُدرة على صناعة القرار، والتَّفسير، والاستنباط، والتَّفكير الإبداعيّ، وغَير ذلك.

وتتفرَّعُ عن الذَّكاء الاصطناعيّ مجالاتٌ معرفيَّةٌ عديدة، منها ذلكَ المجالُ الَّذي نحنُ بِصَدَدِهِ في هذا الكتاب، ونعني «مُعالجَة اللُّغات الطَّبيعيَّة Natural Language Processing»؛ وهو مجالٌ مَعرفيٌّ يُسعى من خلالِهِ إلى توجيه الآلة إلى فهم اللُّغة الطَّبيعيَّة عبرَ مُستوياتها الـمُتعدِّدة، ومُعالجَة وَحداتها في هذه الـمُستويات تحليلًا وتوليدًا، وخَلق بيئةٍ تفاعُليَّةٍ قادرةٍ على تحقيق التَّواصُل بينَ الإنسان والآلة. وحينَ نتحدَّثُ عن اللُّغة الطَّبيعيَّة، فنحنُ مَعنيُّونَ بصُورتَيها الرَّئيسَتَين، المكتوبة والمنطوقة.

واللُّغةُ العربيَّةُ إحدى اللُّغات الطَّبيعيَّة الَّتي تحظى بعناية الباحثينَ في الذَّكاء الاصطناعيّ عُمُومًا ومُعالجَة اللُّغات الطَّبيعيَّة على وجه الخُصُوص. ذلكَ أنَّها واحدةٌ من أكثر اللُّغات انتشارًا في عالمنا الـمُعاصِر؛ حيثُ تأتي في المرتبة الرَّابعة من حيثُ عدد مُستخدميها، بعدَ الصِّينيَّة والأرديَّة-الهنديَّة والإنجليزيَّة.

كذلكَ فإنَّها واحدةٌ من أكثر اللُّغات الطَّبيعيَّة تناميًا في العالَم. ويكفي أن نعرفَ أنَّها انتقلَت من المرتبة السَّادسة في مُنتصَف العقد الأوَّل من القرن الحادي والعشرينَ إلى المرتبة الرَّابعة في مُنتَصَف العِقد الثَّاني، من حيثُ عددُ مُستخدميها الَّذي يقترِبُ من نصف المليار نَسَمة. ولعلَّ هذا الانتشارَ والتَّنامي يُعطينا صُورةً استشرافيَّة عمَّا يُمكنُ أن تصلَ إليه العربيَّةُ في الـمُستقبل القريب.

إنَّنا نُؤمنُ أنَّ اللُّغةَ العربيَّةَ قابلةٌ للـمُعالجَة الآليَّة، ومُهيَّأةٌ للوُصُول إلى مُستوياتٍ مُتقدِّمةٍ للغاية، إذا توافرَت لها المواردُ والطَّاقاتُ من ناحية، وإذا توافرَ لها الدَّعمُ الـمُناسبُ من ناحيةٍ أخرى. ذلكَ أنَّ طبيعةَ العربيَّة قياسيَّةٌ في شَطرٍ كبيرٍ منها، لا سيَّما في بنيتِها الصَّوتيَّة وقوانينها الصَّرفيَّة. وفي الوقت ذاته، فالعربيَّةُ لُغةٌ مُعرَبة، وتتمتَّعُ بنظامٍ اشتقاقيٍّ توليديّ، وتتمتَّعُ أيضًا بنظام كتابيٍّ خاصّ؛ وهي أمورٌ تُمثِّلُ إشكالاتٍ ليسَت هيِّنة؛ ولن تكونَ مُعالجَةُ هذه الإشكالات يسيرةً ما لم تتوافر السُّبُلُ والأدواتُ الـمُؤدِّية إلى ذلك.

ونحنُ نُقدِّمُ هذا الكتابَ (العربيَّة والذَّكاء الاصطناعيّ) الَّذي نسعى من خلالِهِ إلى الوُقُوف على علاقة اللُّغة العربيَّة بالذَّكاء الاصطناعيّ، وما انتَهَت إليه جُهودُ مُعالجَتِها آليًّا، وما يُمكنُ أن يُقدِّمَهُ الذَّكاءُ الاصطناعيُّ للعربيَّة مُستقبلًا. وننشُدُ في عملِنا هذا

الإجابةَ عن أسئلةٍ، منها: ما الذَّكاء الاصطناعيُّ، وما أطوارُهُ، وما إشكالاتُه؟ وما علاقةُ اللُّغات الطَّبيعيَّة بالذَّكاء الاصطناعيّ، وما علاقةُ العربيَّة خُصُوصًا به؟ وإلى أيِّ مدًى يُمكنُ توظيفُ الذَّكاء الاصطناعيّ في مُعالجَة اللُّغة العربيَّة المكتوبة؟ وما واقعُ توظيفِهِ في مُعالجَة اللُّغة العربيَّة المنطوقة؟ وما أبرزُ تحدِّيات حوسَبة العربيَّة؟ وكيفَ يُمكنُ الإفادةُ من تقنيات الذَّكاء الاصطناعيّ وتطبيقاته في تعليم العربيَّة؟

يشتملُ الكتابُ على أربعة فُصُول؛ حيثُ يأتي الفصلُ الأوَّل بعُنوان (الذَّكاء الاصطناعيّ ونَمْذَجة اللُّغات الطَّبيعية: الطُّموح، والواقع، والآفاق)، ويُمثِّلُ تقديمًا مُهمًّا وثريًّا لمفهوم الذَّكاء الاصطناعيّ وإرهاصاتِه وواقعِهِ وآفاقِهِ وعلاقتِه باللُّغات الطَّبيعيَّة، ويُجيبُ الفصلُ أيضًا عن تساؤُلاتٍ مطروحةٍ بشأن طبيعة اللُّغة العربيَّة وقابليَّتِها للـمُعالجَة الآليَّة. ويأتي الفصلُ الثَّاني بعُنوان (الـمُعالجَة الآليَّة للُّغة العربيَّة المكتوبة: مُقدِّمة في ذكاء الآلة)، ويتناولُ رُؤيةً منهجيَّةً حولَ مُستويات اللُّغة العربيَّة المكتوبة ووحداتها والتَّوجيه الآليّ لها ووسائل الـمُعالجَة السَّطحيَّة والعميقة، مع الإشارة إلى تطبيقات مُعالجَتِها آليًّا في كُلِّ مُستوًى على حِدة. أمَّا الفصلُ الثَّالثُ فيأتي بعُنوان (الـمُعالجَة الآليَّة للُّغة العربيَّة المنطوقة)، ويتناولُ طرحًا لـمُكوِّنات الصَّوت اللُّغويّ وآليَّات تحليله حاسُوبيًّا، كما يعرضُ لبعض التِّقنيات الصَّوتيَّة الحاسُوبيَّة وآليَّات بنائِها وتطويرها، ويُركِّزُ خُصُوصًا على تقنية تحويل النَّصّ العربيّ المكتوب إلى صوتٍ منطوق. ويأتي الفصلُ الرَّابعُ أخيرًا بعُنوان (الذَّكاء الاصطناعيّ وتعليم اللُّغة العربيَّة: نحو مِنَصَّةٍ تعليميَّةٍ مُتكامِلة). ويتناولُ هذا الفصلُ رُؤيةً حولَ كيفيَّة الإفادة من الذَّكاء الاصطناعيّ في تعليم مهارات اللُّغة العربيَّة، مع التَّركيز على مهارَتَي الكتابة والتَّحدُّث، ويعرضُ الفصلُ أيضًا مُقتَرَحًا لـمِنصَّةٍ تعليميَّةٍ ذكيَّة، يُمكنُ تطويرُها واستثمارُها في تعليم اللُّغة العربيَّة بمُستوياتِها الـمُختلفة.

وبعد؛

فإنَّنا نرجو أن يكونَ هذا الكتابُ فاتحةً للباحثينَ في ميادين الذَّكاء الاصطناعيّ، خُصُوصًا أولئكَ المعنيِّينَ بالـمُعالجَة الآليَّة للُّغة العربيَّة، لطَرق آفاقٍ أكثرَ سعةً ورحابة من التَّفكير والإبداع، سعيًا إلى الوقوف على إشكالات حوسبة العربيَّة والعمل على

مُعالَجتِها، وتقويم أدواتها وخوارزماتِها ومواردها، والانتقال بها إلى مُجتَمَعٍ مَعرفيٍّ قادرٍ على تطويع الآلة لتحقيق نهضته وبناء ملامح مُستقبلِه.

نسألُ الله تعالى أن يتقبَّلَ هذا الجهدَ بالذِّكر الحَسَن والأجرِ الجزيل، وأن يجعلَهُ من العلم الَّذي ينفعُ أصحابَه بعد مماتهم.

ربَّنا عليكَ توكَّلنا وإليكَ أنَبنا وإليكَ المصِير.

المُحرِّر

د. المُعتزّ بالله السَّعيد

الفصل الأول

الذَّكاءُ الاصطناعيُّ ونَمْذَجةُ اللُّغاتِ الطَّبيعية

الطموح، والواقع، والآفاق

د. محمد عطية محمد العربي أحمد

ملخَّص

يملأ الحديثُ عن مصطلح «الذكاء الاصطناعي» وأخبارُ تقنياته والتبشيرُ بإنجازاته في السنوات الأخيرة وسائلَ الإعلام في السنوات الأخيرة، فما هو الذكاء الاصطناعي؟ من أين وكيف نشأ؟ وما تاريخه؟ ما هي عَثَراتُه وما هي نجاحاتُه؟ هذا ما يُفَصِّلُهُ القسمُ الأولُ بأجزائه الثمانية التي تمثِّلُ الأساسَ الذي تَنبْنِي عليه الأقسامُ التالية من الفصل. وينتقل القسم الثاني إلى التعرف على كيفية معالجة المسائلِ المعتبَرَةِ في إطار الذكاء الاصطناعي بمدرسَتَيْه؛ التحليليةِ، والعدديةِ الإحصائية.

وينصَبُّ التركيز في القسم الثالث على إحدى المسائل الكبرى المطروحة في مجال الذكاء الاصطناعي - إن لم تكن أكبرها على الإطلاق - ألا وهي نَمْذَجة اللغات الطبيعية ومعالَجَتُها آليًّا والتحديات الأساسية التي تواجهُها. وفي القسم التالي له يقف القارئ على الوضع الراهن للبحث العلمي والتطور التقني للمعالَجَة الآلية لِلُّغات الطبيعية.

ثم يناقش القسم الخامس ما إذا كان تطور المعالجة الحاسوبية الآلية للغات الطبيعية سيمضي إلى أن يحاكي مَلَكات الإنسان العاقل وربما يتجاوزه، أم أنَّ هناكَ سقفًا لذلك لا يمكن المروق منه. وهي ليست فقط قضية فلسفية نظرية عميقة، بل إنها محدِّدة لمسار البحث والتطوير على هذا الصعيد.

وينظر القسم السادس من الفصل في حال اللغة العربية مع الذكاء الاصطناعي من حيث ما إذا كانت لها خصوصية عن غيرها من اللغات الطبيعية سواء في نمذجتها أو معالجتها الآلية، وما تلك الخصوصياتُ إن وُجِدَتْ؟ أما القسمُ السابع في مُخْتَتم الفصل فإنه يستشرف في ضوء ما سبق من أقسامه آفاقَ مستقبل اللغة العربية مع الذكاء الاصطناعي على رجاء أن يكون هذا الفصل الأول كُلُّه زادًا للقارئ وعَوْنًا له على استيعاب بقية فصول الكتاب والتفاعل الإيجابي معها.

الكلمات المفتاحية:

التعلُّم الحاسوبي، الذكاء الاصطناعي، اللُّغات الطبيعية، اللُّغة العربية، النَّمْذَجة الحاسوبية، المقارَبة التحليلية.

١. تاريخٌ مختصَرٌ للذَّكاءِ الاصطناعيّ.

تعلِّمُنا مدرسةُ التاريخ على اختلاف مواضيعها أن جذورَ النشأة دائمًا ما تطبع أثَرَها على الحاضر وعلى المآلات، والذكاءُ الاصطناعيُّ ليس بِدْعًا مما دوَّنَه التاريخ، ولذلك فليس فَضْلةً ولا تَزَيُّدًا أبدًا الاستفاضةُ في القسم الافتتاحي من هذا الفصل من أجل وضع القارئ في صورة نشأة الذكاء الاصطناعي وتطوره التاريخي ومبادئ آليات عَمَلِه الأساسية بتبسيطٍ غير مُخِلٍّ، ومن ثَمَّ علاقَتِهِ بمَيْكَنة اللغات الطبيعية. ولعل القارئَ يجني بعد مطالَعَتِهِ لهذا القِسْمِ الأول ثمارَ ذلك أثناء دراستِهِ للأقسام التالية من الفصل؛ استيعابًا سَلِسًا لما سيَرِدُ فيها من اصطلاحاتٍ ومسائلِ حوسبة اللُّغات الطبيعية، وفهمًا لماذا اتُّخِذَتْ بعضُ الخيارات دون غيرها في مساراتِها، وربطًا بما يرد في ختام الفصل من استشرافات، ولعله يحصِّلُ عِلاوةً على ذلك كله متعةَ تتبُّع قصة ميلاد علمٍ جديدٍ من رحم تفاعلاتٍ مطوَّلةٍ محتدِمةٍ بين ما كان قائمًا قَبْلَه من فلسفةٍ وعلومٍ وتقنياتٍ معًا.

١, ١ حُلمٌ قديمٌ وسعيٌ مستمرٌّ

ما بَرِحَ كِبارُ الحكام والقادة والمخطِّطين في مختلف أنحاء العالم عبر أزمنة التاريخ قديمِها وحديثِها تداعبُ خواطرَهم أحلامُ امتلاكِ جيوشٍ كبيرةِ العدد من العُمَّال أو المقاتلين غير البشريين المسخَّرين تمامًا للقيام بما يقوم به البشر من وظائف العمل أو القتال ولكن دون تكبُّد أعباء البشر الحقيقيين الذين يحتاجون للراحة وللطعام وللكساء وللعلاج وللمأوى وللترفيه، وللتعليم والتدريب، وكذلك لمراعاة معتقداتهم ومعالجة شؤونهم المعنوية، ومع ذلك كُلِّهِ فقد يتمردون ويخرجون عن السيطرة. فمن يمتلك دون خصومِهِ مثلَ تلك الجيوش من الآليِّين أشباه البشريين تحت كامل سيطرته فلا شك أنه سيهزمهم اقتصاديًّا أو يسودُهم عسكريًّا.

فقبل حوالي ثلاثة آلاف سنة في الصين يُحْكَى أن مهندسًا يُدْعَى «يان شِي» قدم للإمبراطور آنذاك «تشاو مِيُو وانج» آلة ميكانيكية من صنعه على هيئة آدمي وبحجمه الطبيعي تستطيع الخطو وتحريك اليدين وتأتي ببعض تعبيرات الوجه، وكذلك يُذْكَر الرومانيُّ «هِيرون» عالمُ الرياضيات والمهندسُ الميكانيكي الشهير الذي عاش في مدينة الإسكندرية في مصر قبل حوالي أَلْفَيْ سنة والذي اشتهر بآلاته التي ميكَنت العديد من الوظائف التي اعتاد البشر القيام بها؛ مثل العزف على آلة «الأورغُن» وبرمجة المؤثرات السمعية والبصرية على المسرح، وتماثيل ميكانيكية على هيئة بشرية تصب الشراب وأخرى تقوم بفتح وغلق أبواب القصور والمعابد.

ومن بين مئات الآلات التي صممها ونفذها قبل نحو ثمانية قرونٍ أحدُ رُوَّادِ الهندسة الميكانيكية ومشاهيرِها «بديع الزمان أبو العز ابن إسماعيل الجَزَريّ» - الذي خدم في بلاط «بَني آرتوق» في «دِيار بَكر» ثم في البلاط الأيوبي - عشراتُ الآلات الميكانيكية على هيئةٍ بشريةٍ تقوم بمهامَّ متنوعةٍ؛ منها مثلاً خادم آلي يصب الشراب، وآخر يصب الماء للوضوء وغسيل اليدين، وجَوْقة آليين تعزف الآلات الموسيقية تلقائيًّا، وغير ذلك كثير. ثم نجد على هذا الطريق في نهاية القرن الثامن عشر الميلادي عددًا من الأوروبيين نذكر من بينهم صانع الساعات البارع الفرنسي «بِيير جاك درو» وابنه «هنري لُوِي» الذي اشتهرت دُمَاهُما الميكانيكية ذات الهيئة البشرية المتقَنة بين دُمْيةٍ تعزف الموسيقى وأخرى تخط حروفًا بالقلم ونحو ذلك، وقد تميزت هذه الدُمَى الميكانيكية بوجود ذاكرة ميكانيكية تُمثِّل «برنامجًا» (خطوات حركية متسلسلةٍ معرَّفةٍ سلفًا) ولا تزال ثلاثة من هذه الدُمى معروضةً في متحف الفن والتاريخ بمدينة «نوشاتيل» في «سويسرا».

ويحفل تاريخُ مختلف الأمم عتيقُه وقريبُه مثلما يحفل التراث العريض لأدب الخيال العلمي بالطرائف والنفائس من أمثلة تلك الآلات التي حاولت محاكاة البشر في ملمحٍ أو أكثر، وتشير كلها دون لبس إلى قِدَم وتواصُل الولع بحلم صناعة «أشباه البشر». وكثيرًا ما يستخدم العامَّةُ اصطلاح « الخادم الآلي robot»[1] بدلًا من «شَبِيه البشر humanoid»؛ وخصوصًا في سياق التحكُّم والتسخير، والواقع أن الاصطلاحين ليسا متطابقين كما سيرد في القسم الفرعيّ ١,٥ من الفصل.

١ - كلمة « رُوبُوت» مشتقة من كلمة « رُوبُوتا robota» باللغة التشيكية التي تعني «السُّخْرة».

٢,١ . نهضةُ الرِّياضياتِ قاطرةً للفيزياء والتِّقنية.

على صعيدٍ آخرَ وثيقِ الصِّلةِ؛ فإن حقل الرياضيات قد انتقل إلى حقبةٍ جديدةٍ مع مطالع القرن السادس عشر الميلادي حيث نقل الأوروبيون معارف الرياضيات من العربية إلى اللاتينية واتسعت وَقْتَها كثيرًا قاعدة المشتغلين بها انطلاقًا من الإمارات الإيطالية، ويشير العديدُ من مؤرِّخي العلوم إلى عام ١٥١٩م - حين توصل الإيطاليون إلى صيغة الحل لإيجاد جذور «كثيرة الحدود polynomial» من الدرجة الثالثة - على أنه نقطة انتقال التراكم المعرفي في الرياضيات من تراكمٍ رَتيبٍ بَطِيءٍ قبلها إلى تراكُمٍ على منحنًى متزايد التسارع بعدَها فاستقر التعبير الرمزي في الرياضيات، وصيغت نظرية عامة للأعداد، وللجبر، ثم ابتُكِرَ حساب التفاضل والتكامل، والمعادلاتُ التفاضليةُ ودراسةُ حلولها، والإحصاءُ وحسابُ الاحتمالات، والجبرُ الخَطِّيُّ، وحسابُ المصفوفات، والأعدادُ المركَّبة، ثم هندسةُ السطــوح المنحنية (الهنــدسة اللا إقْلِيدِيَّة Non-Euclidean geomerty) ... وغير ذلك كثيرٌ مما يضيقُ المقامُ هنا عن ذكره.

ويعني ذلك كله امتلاكَ العديد من الأدوات الناجعة للتمكين من الرصد والتحليل والتصميم والحساب والإدارة والتحكم ، فكانت هذه الأدوات أعمدةً قامت عليها نهضة أخرى هائلة في اكتشاف وتطور علوم الطبيعة (الفيزياء[١])، والتي بدورها كانت أساسًا آخَرَ قامت عليه نهضةٌ تقنيةٌ صناعيةٌ كبرى بدءًا من النصف الثاني للقرن الثامن عشر الميلادي تتابعت موجاتها لتبلغ بنهاية الحرب العالمية الثانية نضجًا وتمكينًا كبيرًا، لا سيما على محاور الميكانيكا، والبصريات، والطاقة، ثم الكهرومغناطيسيات، والاتصالات، وترميز البيانات وتشفيرها.

ومن المفيد هنا الإلماحُ إلى أن الرياضيات قد انتقلت عبر ذلك الوقت بالتدريج من مجرد تداول ومعالجة المعادلات والتساويات والمتباينات - مهما كان تعقيدُها - إلى القدرة على تقديم حلولٍ متكاملة لمسائلَ مُرَكَّبةٍ مصوغةً على هيئة إجراءات عديدة الخطوات لمعالجَة مُدخَلات المسألة تجمع إلى جانب المعادلات حدودًا شرطيةً تعرِّف الانتقال بين هذه الخطوات مما يعرِّف ضمنيًّا المسارات (السيناريوهات) الممكِنة لحلول

١- وتُلْحَق بها الكيمياءُ، وتطويرُ المواد.

المسألة قيد المعالجة، وهو ما اصطُلِحَ على تسميته بـ «الخُوارِزم[1] algorithm» وهو الصورة الرياضية التي يمكن تنفيذها أو ميكنتها عبر آلةٍ مناسبةٍ قابلةٍ للبرمجة.

٣,١. انبعاثُ الفلسفةِ وتفاعلُها مع الرياضيات سعيًا وراء قوانينَ للتفكير البشري.

وعلى صعيد ثالثٍ يتكامل مع سابقَيْهِ فقد اقتفى مسارُ الفلسفة خُطًا مشابهةً حين بدأ انبعاثها في أوروبا خروجًا عن الإطار المعرفي الجامد الذي فرضته الكنيسة الكاثوليكية لقرون طويلة من ناحيةٍ وحين اكتمل انتقالُ التراث الفلسفي من العربية إلى اللاتينية من ناحيةٍ أخرى، فبزغت في أوروبا أسماءٌ كبيرةٌ في دنيا الفلسفة من القرن السادس عشر فصاعدًا من أمثال «فرانسيس بيكون» إلى «رينيه ديكارت» إلى «جوتْفْرِيد لايْبِينْز» إلى «إيمانويل كانط» ... وغيرهم الذين طرحوا باستفاضة قضايا كبيرة تتصل بماهية المادة والطبيعة، والعلوم، والعقل والإدراك، وقوانين التفكير.

وفي وقت باكرٍ من تلك الحقبة الجديدة كان الاتصال لا يزال وثيقًا بين الرياضيات وعلوم الطبيعة (الفيزياء) والفلسفة، حتى أن الفيزياء كانت تسمى في وقت «إسحاق نيوتن» ومعاصرِيه بـ «الفلسفة الطبيعية»، ولذلك أسفرت الجدلية بين الفلسفة ومعارف الطبيعة والرياضيات متسارعة النمو في القرون من السادس عشر إلى التاسع عشر الميلادية عن السعي الحثيث لفهم قوانين التفكير البشري وتقعيدها بل الطموحِ إلى مَيْكَنة هذا التفكير، والذي أنتج بدوره إنجازاتٍ فارقةً.

ففي منتصف القرن التاسع عشر استطاع الإنغليزي «جورج جون بُوول» صياغة منظومة «المنطق الرياضي الثنائي Boolean Algebra» في صورة قوانين جبرية محكَمة[2]، كما انتشرت في القرن نفسه بين الفلاسفة والـمَناطِقة الرياضيين نزعة لتأسيس

١- وذلك اعترافًا بفضل الفَلَكِيّ وعالم الرياضيات الأبرز في العصور الوسيطة «محمد ابن موسى الخُوَارِزْمِيّ» المتوفَّى في بغداد عام ٨٥٠م، وتحمل كذلك اسمَهُ العملية الرياضية الشهيرة «لوغارِيتْم Logarithm» التي هي معكوس رفعِ عددٍ لأُسٍّ. وهو صاحب كتاب «الجَبْر والمقابَلة»، وهو كذلك من أدخل الصفر في نظام العد العشري. ولا يزال علم الجبر يحمل نفس الاسم العربي في كل لغات العالم «Algebra» ولا يزال كذلك اسم الصفر في مختلف لغات العالم مشتقًا من اسمه العربي؛ فلا رياضياتٍ حديثةً بدون نظام العد العشري ولا بدون الصفر ولا بدون علم الجبر، والخوارزمي رحمه الله هو الرائد المبرَّز في ذلك كله.

٢- وهو الأساس الذي تعمل وفقَه كل الدوائر الإلكترونية التي تنفذ عمليات التحكم والحساب في كل الحواسيب الرقمية بمختلف أنواعها.

الرياضيات بالكامل على أساس مُحْكَمٍ من المنطق البَحْت الذي لا يتضمن سوى مجموعة محدودة من المسلَّمات مع القواعد الاستنباطية الأساسية فقط لا غير، فإذا تم ذلك فلربما صار الطريق مفتوحًا لميكنة عملية البَرْهَنة الرياضية، فكَرَّسَ الألماني «غُوتْلِيبْ فِرِيغِيهْ» حياته المهنية في الربع الأخير من نفس القرن لمحاولة تأسيس الحساب العددي بإحكام على المنطق وأصدر عمله في مجلدات ثلاثة، وقبيل طباعتها اكتشف أحد كبار الـمَناطِقة الرياضيين الآخرين الإنغليزي «بِرْتِرانْد راسِل» تناقضًا يخلخل بنية عمل «فِرِيغِيهْ»، وكان «راسِل» قد أخذ على عاتقه هو القيامَ بهذا التأسيس فنشر بين عامَيْ ١٩١٠م و ١٩١٢م مع أستاذه الفيلسوف وعالم الرياضيات «آلفريد نورث وايِتْ هِيد» عملاً نظريًا ضخمًا آخر[1] يؤسس للحساب العددي على أساس منطقي بَحْتٍ بما يتجنب ثغرة « فِرِيغِيهْ».

وأثناء ذلك فقد كان مجتمع كبار علماء الرياضيات بعد الحرب العالمية الأولى – وقد تشعبت فروعُها وتطاوَلَ بنيانها المعرفي ليبلغ ذُرًا شاهقةً – يمور بالجدل حول ما إذا كانت أُسُسُها متينةً بما يكفي من أجل ضمان أن تحقق منظومة الرياضيات شرطين أساسيين (كان غالب الظن وقتها أنهما متحققان تلقائيًّا) ألا وهما: خلوها من إمكانية تَوَلُّد متناقضاتٍ، والاكتمال (أي قدرتها من حيث المبدأ على برهنة أية حقيقة رياضية يُمْكِنُ اشتقاقُها من مُسلَّماتها الأولية). ولذلك دعا العَلَمُ الأَبْرَزُ مسموع الكلمة في ذلك الوقت الألمانيُّ «ديفيد هِيلْبيرت» إلى مؤتمرٍ عالميٍّ جامعٍ للبت في هذه القضايا. وكانت ثمرةُ ذلك في بداية ثلاثينيات القرن العشرين هي المفاجأةُ الكبرى التي فجرها العبقريُّ النمساويُّ (كِيرْت غِيدِل) بمُبَرْهَنَتَيْهِ البديعتين الشهيرتين عن استحالة اكتمال أيةِ منظومةِ رياضياتٍ وخُلُوِّها من التناقض في الوقت ذاته، وبذلك حَدَّدَ بالبرهان الرياضي سقفًا للرياضيات وأثبت دُونَ شَكٍّ وجودَ حدودٍ للمعرفة فيها مع تبيان أسباب ذلك. ومن رحم تلك الجدلية المحتدمة بين من حاولوا تأسيس الرياضيات على المنطق الرياضي الصافي – ومن ثم إمكانيةَ ميكنة عملية البرهنة واشتقاق الحقائق الرياضية من مسلماتها الأولية – وبين من أثبت رياضيًّا استحالة تحقيق ذلك بالكامل، تشكلت الملامح الأولية لـ «نظرية تعيين الصعوبة Complexity theory» التي واصل رُوَّادٌ مثل الأمريكيّ

---

[1] – ونُشِرَ تحت عنوان «Principia Mathematica» ويعني باللاتينية «المبادئ الأساسية للرياضيات».

المـجَريّ «جون فون نُويْمان» والبريطاني «آلان تيورِنغ» تأسيسها والتي صارت بمثابة اللُّبِّ النظري لـ «علوم الحاسب Computer Science»[1].

٤ , ١ . وبدأ عصرُ الحواسيب الرقمية.

بحلول أربعينيات القرن العشرين الميلادي سواء أثناء الحرب العالمية الثانية وبُعَيدَها، كانت كل الظروف قد اجتمعت من أجل بناء حواسيب رقمية ذات قدرات حسابية عالية تتصف بعمومية الاستخدام؛ بمعنى قابليتها للبرمجة كي تنفذ خوارزمات مختلفة، مع القدرة على استقبالِ بياناتٍ رقميةٍ كمدخَلاتٍ تعمل عليها تلك الخوارِزمات، وكذلك تصدير بياناتٍ رقميةٍ كمخرَجاتٍ لعمل تلك الخوارزمات. وتمثلت تلك الظروف في أن:

- حلم صناعة شبيه الإنسان قائم لم يغب منذ القِدَم، وميكنة التفكير البشري في قلبه، وميكنة القدرات الحسابية حجر أساس فيه.
- الفِكْر الرياضي وأدواته صارت ناضجةً لذلك.
- التقنية الكهرومغنطيسية تُمَكِّنُ من ذلك.
- الصراع العسكري العالمي المشتعل والسعي وراء امتلاك كل أسباب القوة فيه يتوجان ويوظفان كل ما تَقَدَّمَ من ظروف.

ولذلك كان من الطبيعي أن تُبْنَى أثناء ذلك «الحواسيب الرقمية» الأولى التي يمكن برمجتُها، وكان من طبائع الأمور أيضًا أن يتصدر أكبرُ المنتصرين في الحرب العالمية

١- هناك اعتقاد غير صحيح أن «علوم الحاسب» لا وجود لها إلا بوجود «الحواسيب الرقمية»، والصواب أن علوم الحاسب بمختلف أدواتها ذات وجود سابق ومستقل عن تلك الحواسيب، ولا تفترض لعملها إلا توافر «وحدات» تقوم بإجراء العمليات الحسابية والمنطقية الأولية؛ فربما تكون هذه الوحدات آلات إلكترونية أو كهربائية أو حتى ميكانيكية بل ربما تكون بشريةً؛ ومن الطريف في هذا المقام الإشارة إلى أن «وكالة الطيران والفضاء الوطنية: ناسا» الأمريكية الشهيرة كانت منذ تأسيسها في نهاية الخمسينيات من القرن الميلادي الماضي حتى منتصف سبعينياته توظف كتائب من الموظفين المتخصصين في الرياضيات العددية الذين كانوا يعملون كحواسيبَ بشريةٍ لمساعدة علماء الفلك وعلماء الصواريخ وعلماء الطيران حيث تقوم تلك الكتائب بتنفيذ خوارزمات للحصول على حلول عددية لمعادلات معقدة يحتاجها هؤلاء العلماء في تصاميمهم ويتعذر الحصول على حلول تامة مغلقة لها.

الثانية «الولاياتُ المتحدةُ الأمريكيةُ» ريادةَ هذا المضمار[1]، كما لم يكن مستغربًا أن تكون الاستخدامات الأولى لتلك الحواسيب استخداماتٍ عسكريةً وأن تعتبر هي نفسها أسلحةً حربيةً ضاربة[2].

وعلى الرغم من أن الوحدات الحسابية الأولية في تلك الحواسيب كانت ذات تقنيةٍ كهربائيةٍ ولم تتحول إلى إلكترونية بالكامل إلا بعد عقدين من الزمان فإن «مِعْمارَها architecture» لا يزال في جوهره هو المتبع في بناء الغالبية العظمى للحواسيب التي نستخدمها حتى يومنا هذا. والقصةُ بعد ذلك من تحول الحواسيب للتقنية الإلكترونية بدلاً من الكهربائية ثم تضاؤل هذه الوحدات الأولية الإلكترونية وزيادة كثافتها في وحدة المساحة بنمطٍ أُسِّيٍّ، ومن ثم تضاعف القدرات الحسابية والتخزينية للحواسيب كل حوالي ثمانية عشر شهرًا (وهو ما يعرف بقانون «مُوورْ»)، فضلاً عن التطور المطرد في وسائل إدخال وتصدير وعرض البيانات، واندماج الحواسيب مع تقنيات الاتصالات والشبكات الرقمية ... إلخ إلخ حتى نصل إلى وقتنا الحاضر هي قصةٌ معروفةٌ نعيشها ولا حاجة بنا للاستفاضة فيها في هذا المقام.

٥, ١. إذن، ما هو الذكاء الاصطناعي بالضبط؟ وما هو هدفه النهائي؟

بعد انتشار أخبار صناعة الحواسيب الرقمية القابلة للبرمجة وإسهامها في حل مسائل حسابية كبيرة، اشتعلت الحماسة وتجددت الآمال بين الأوساط العلمية في اقتراب تحقيق حلم العُصور الذي طال انتظاره بصناعة «أشباهٍ للبشر» ذوي قدراتٍ ذهنية وحركية وانفعالية تشابه ما للإنسان من قدرات بما يتجاوز الدُمَى الميكانيكيةَ القديمة التي أشرنا إليها في القسم الفرعي ١,١ من هذا الفصل. فقد كانت الفلسفة المادية الجامحة هي

١- يبرز في ملحمة مشروعات بناء أول تلك الحواسيب اسم عالم الرياضيات والفيزياء النافذ «جون فُونْ نويْمان» مصمِّمًا أول ومديرًا لتلك المشروعات، إلا أن الأقل شهرةً والأسبق ريادةً هو المهندس الألماني «كونْراد زُوسِهْ» الذي كانت حاسباته الأولى ميكانيكيةً كهربيةً مختلطةً، ولعل هزيمة ألمانيا ومحورها في الحرب أمام الولايات المتحدة الأمريكية وحلفها هو ما أخفت ذِكْرَه.

٢- استُخْدِمَت تلك الحواسيبُ أثناء الحرب العالمية الثانية بكثافةٍ كأدواتٍ لفَكِّ شفرة الرسائل العسكرية السرية، كما ساهمت بفعاليةٍ وقتذاكَ في مشروع «مانْهاتِنْ» لصناعة القنابل الذرية «الانشطارية» الأمريكية التي ألقيت على «هيروشيما» و«ناجازاكي» اليابانيتين في نهاية الحرب، ثم كذلك في تصميم القنابل الهيدروجينية «الاندماجية» الأشد فتكًا ضمن سباق التسلح المحتدم مع الاتحاد السوفْيِيتِّي وحلفه الشيوعي والتي فَجَّرَت الولايات المتحدة الأمريكية أُولاها تجريبيًّا بنجاح عام ١٩٥٢م.

السائدة بين الأوساط العلمية في حقبة الأربعينيات والخمسينيات من القرن العشرين الميلادي، وكان مُنْطَلَق الحماس أن تلك الحواسيب الرقمية يمكنها أن تحاكي دور الدماغ البشري الذي هو من وجهة نظر تلك الفلسفة المادية في النهاية مجرد آلة حسابية منطقية متطورة، ومع زيادة القدرات الحاسوبية الرقمية تدريجيًّا فإنها سوف تنافس الدماغ البشري وربما تَفُوقُه! كما أن الهندسة الميكانيكية متقدمةٌ (وقتذاك) بما يمكِّنها من محاكاة الجسد بعد ربطها كهربيًّا والتحكم فيها بواسطة الدماغ (الحواسيب الرقمية).

بعد وقت ليس بالطويل من المحاولة والبحث المعمَّق للمسألة استبانت الصعوبة البالغة لصناعة «شبيهٍ للبشر humanoid» حتى ولو بقدرات رضيع حديث الولادة؛ فالحاسبات الرقمية كانت وقتها تحتل أجنحةً كاملةً من مباني كبيرة، وتستهلك طاقةً كهربيةً هائلةً، ومع ذلك فإن قدراتها الحسابية عند توظيفها لتنفيذ خوارزمات «البحث الشجري tree search» [1] من أجل المفاضلة بين البدائل واتخاذ القرار تجعل تشغيل تلك الخوارزمات بطيئًا جدًّا؛ وبالتالي فإن استجابة ذلك «الشبيه» لمثيرٍ خارجيٍّ ما لن تتبدى إلا بعد ساعاتٍ من حصوله. كما كان الجمع بين متانة الجسد واستدامته ومرونته وسلاسة حركته، والحفاظ على اتزانه، وتوفير استقلاله من حيث الطاقة وتجددها ... إلخ كلها أمورًا متعذرةً تمامًا على الصعيد الميكانيكي. وكذلك كانت وسائل اكتساب البيانات من البيئة المحيطة ضوئيًا أو صوتيًا أو تَلَمُّسًا أضعف كثيرًا من أن تشابه البصر أو السمع أو الإحساس عند البشر، ناهيك عن تفسيرها وتحليلها للخُلوص إلى معلومةٍ مفيدة. وعلاوةً على ذلك كله فلم تكن هناك آليات يُعوَّل عليها لاكتساب الخبرات وللتعلُّم التلقائي وهو ملمح جَوْهريٌّ يميز البشر عن الآلات.

ثم ماذا عن القدرات اللغوية لذلك «الشبيه البشري» وهي وسيلة التواصل التي لا بد منها كي تُقنع من سيستخدمه أو يتعامل معه أنه «بشري»؟ لم يكن لدى أحد وقتها أية فكرة عن ميكنة اللغة تتجاوز عملية الترميز[2].

نجمَت عن إدراك هذه العقبات كلها في ذلك الوقت الباكر عدةُ استجاباتٍ تهمنا أحدها في سياق هذا الفصل؛ ذلك أن العلماء المهتمين أيقنوا أن معالجة تحدي صناعة

١- سوف نأتي في القسم الفرعي التالي من هذا الفصل بشيء من التفصيل على شرح البحث الشجري.

٢- وذلك استلهامًا من إنجازات «نظرية الاتصال المعلوماتي Information theory» التي كانت قد تأسست على يَدِ الأمريكي «كلود شانون» ثم استوت على عودها وقتذاك قبل حوالي عقد من الزمان.

«شبيه البشر» جُمْلَةً واحدةً يكاد يقع في مقام الاستحالات، ولذلك لَجَؤُوا إلى «المعالَجة بالتجزئة divide and conquer» وهي الاستراتيجيةُ المجرَّبة نجاعتُها مع أمثال تلك التحديات الكبرى. ومن وقتها استقر تقسيم العمل في تطوير أشباه البشر الآليين إلى الجناحين التاليين:

- «الذكاء الاصطناعي Artificial Intelligence» ويشتهر كذلك بالاختصار «AI» ويُعْنَى إجمالاً بمحاكاة وظائف «المُخّ» في رأس هذا الآلي.
- «الرُوبُوتِيَّات Robotics» وتُعنَى إجمالاً بمحاكاة وظائف «باقي الجسد» في هذا الآلي.

ويطبق الجدول التالي هذا التقسيم على المسائل التي شكلت تحديات كبرى أمام صناعة «أشباه البشر» مما ذكرناه آنفًا في هذا القسم الفرعي:

مسائل الذكاء الاصطناعي (في الرأس)	مسائل الروبوتيات (في باقي الجسد)
اكتساب البيانات من البيئة المحيطة ضوئيًّا أو صوتيًّا	اكتساب البيانات من البيئة المحيطة تَلَمُّسًا
تحليل البيانات المكتسَبة للخُلوصِ إلى معلوماتٍ مفيدة	الموازنة بين متانة الجسد وخفة حركته ومرونته
المفاضلة بين البدائل واتخاذ القرار	صيانة الجسد واستدامته
إدارة الذاكرة والاستدعاء الموضوعي لمحفوظاتها	تخطيط حركة الأطراف وتحيينها لحظةً بلحظة
تحقيق التراكم المعرفي	سلاسة حركة الجسد والحفاظ على اتزانه
التعلُّم التلقائي	استقلال الجسد الآلي في اكتساب الطاقة وتجددها
القدرات اللغوية تحليلاً وبِناءً	الدفاع عن الجسد في مواجهة الأخطار الخارجية
مسائل وتحديات أخرى كثيرة ...	مسائل وتحديات أخرى كثيرة ...

الجدول ١: تصنيف أَوَّلِيٌّ لمسائل صناعة «شبيه البشر» بين حَقْلَيْ الذكاء الاصطناعي والرُّوبُوتِيَّات

وفيما نستأنف رحلتنا في هذا الفصل مع الذكاء الاصطناعي مُضيِّقِينَ بؤرةَ اهتمامنا أكثر لتركز على معالجة الذكاء الاصطناعي لِلُّغات الطبيعية وخصوصاً العربية، فمن

المفيد ونحن نودع «الروبوتيات» أن نُلْمِحَ إلى أنها قد حققت الآن بعض النجاحات المتقدمة من حيث المواد التي تُصْنَع منها وتحقيق المتانة وخفة الوزن معًا، وكذلك من حيث تخطيط وتنفيذ الحركة؛ فلم تعد «الروبوتات» في زمننا هذا متقيدةً حَصْرًا بالشكل البشري للجسد؛ فهناك الآن الكثير والكثير منها تركِّز على تحقيق كفاءة الوظائف الحركية حتى ولو استلهمت ميكانيكيَّةَ وهيئةَ كائنات أخرى كالحشرات مثلاً، أو كأجزاء فقط من الجسد كالذراع مثلاً، أو حتى ككائنات مجهرية دقيقة.

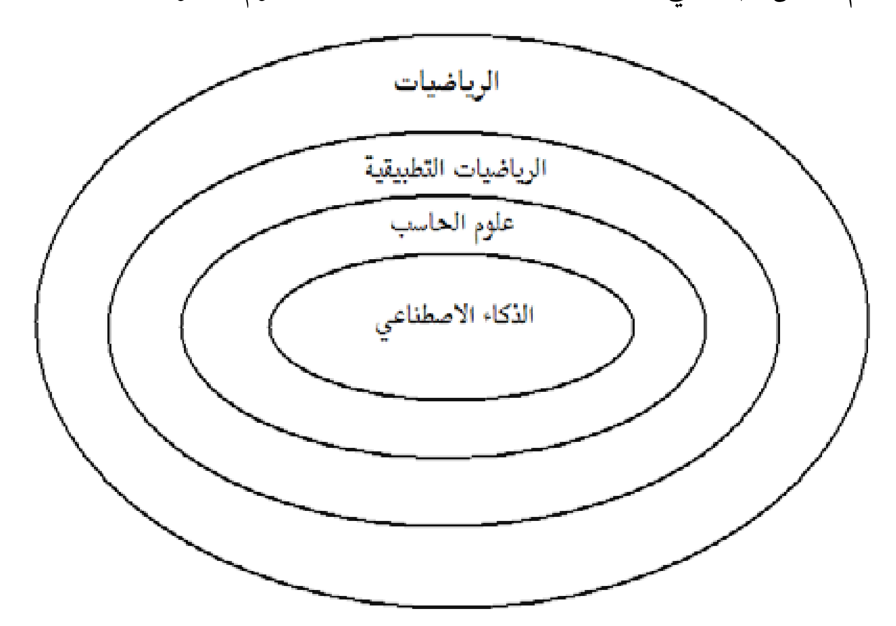
ورغم أن كثيرًا من المسائل في «الروبوتيَّات» ما زالت تمثل تحدياتٍ كبيرةً أمام البحث العلمي والتطوير الهندسي، إلا أن «الروبوتات» قد استقرت الآن كمكون فائق التطور في بنية أنشطةٍ وصناعاتٍ أساسية مثل صناعة السيارات والطائرات والإلكترونيات والطب واستكشاف الفضاء الخارجي والتعدين ونزع الألغام، وبجدوى اقتصادية مرتفعة للغاية. وقبلَ مواصلة رحلتنا مع «الذكاء الاصطناعي» لا بد لنا من تقديم تعريفٍ لهذا المصطلَح الذي سُكَّ للمرة الأولى عام ١٩٥٦م واستقر من وقتها في مختلَف الأدبيات العلمية والإعلامية والأدبية، وفيما يلي محاولة لصياغة مثل هذا التعريف:

- الذكاء الاصطناعي هو العلم الذي يشتغل بابتكار وتطوير خُوارِزْمات مفيدة تسهم في المحاكاة الآلية لقُدُرات الدماغ البشري؛ من إدراكٍ للبيئة المحيطة، والاستجابة المناسبة لمثيراتها، وتعلُّم، وتخطيطٍ، وإيجادٍ لحلولٍ للمسائل المستجدة، والتواصل اللغوي، وإدارةٍ للتراكم المعرفي، ... إلخ (ويُطْلِقُ البعض على هذه القُدُراتِ وأمثالها «الـمَلَكات العليا» للإنسان) ويَخْرُجُ من هذا التعريفِ المسائلُ المعلومُ لها تعريفٌ رياضيٌّ مُحْكَمٌ والمعلومُ لها حلولٌ مفيدةٌ مُبَرْهَنةٌ رياضيًّا.

- من أجل أن يكون لأي حل/ خوارزم وجود مفيد في علوم الحاسب، لا بد ألا تتخطى درجة «صعوبته complexity» على مقياس «نظرية تعيين الصعوبة Complexity theory» عتبةً معينةً؛ وإلا فلن يكون مفيدًا حيث لا يمكن عندئذٍ تنفيذُه على الحواسيب الرقمية. فعلى سبيل المثال؛ لو تزايدت العمليات الحسابية و/ أو الذاكرة المطلوبة لتنفيذ خوارزمٍ ما (لتقديم خدمةٍ ما لعدد من المشتركين فيها مثلا) بنمطٍ أُسِّيٍّ مع زيادة حجمِ المسألة (عدد المشتركين) فإن درجة صعوبته الأُسِّيَّةَ تلك تجعل تنفيذَه حاسوبيًّا أمرًا مستحيلاً مع نمو عدد المشتركين وبالتالي تنعدم إفادته.

ومن المهم للغاية ملاحظة أن هذا التعريف تعريفٌ متحركٌ في الزمن؛ حيث يمكن وفقًا له أن تقع مسألة ما ضمن اهتمام الذكاء الاصطناعي في وقتٍ ما ثم تخرج منه لاحقًا إذا اكتُشِفَ لها تعريفٌ رياضيٌ محكَمٌ وحلٌّ مفيدٌ مبرهَنٌ رياضيًّا. ومن الأمثلة على ذلك تصميم «لغاتٍ عالية المستوى High-Level Languages» لبرمجة الحواسيب الرقمية وبناء «مترجِماتٍ compilers» آليةٍ لها.[1]

ومن المهم كذلك في ختام هذا القسم الفرعي أن نُعَيِّنَ موضع «الذكاء الاصطناعي» على خارطة العلوم بالتعرف على سلسلة نَسَبِهِ فيها - انظر (الشكل ١) أدناه - فالذكاء الاصطناعي ينتمي لحقل «علوم الحاسب» وهو فرعٌ متقدِّمٌ من «الرياضيات التطبيقية» وهي بدورها جزء من «الرياضيات». وبطبيعة الحال فإن هذا النَّسَبَ لا يمنع الذكاء الاصطناعي من استلهام بعض الأفكار أو المبادئ من علوم أخرى مثل «الفيزياء»، أو «علم النفس الإدراكي cognitive pscychology» أو «العلوم التربوية».



الشكل ١: نَسَبُ «الذكاء الاصطناعي» بين العلوم

---

١ - وبسبب ارتباطها الوثيق بتطور «نمذجة اللُّغات الطبيعية Natural Language Modeling» فسوف نتوقف مع هذه المسألة المضروب بها المثال هنا بمزيدٍ من التفصيل في القسمين الثاني والثالث من هذا الفصل.

٦,١. نجاحاتٌ مبكِّرةٌ وتفاؤلٌ عريض

كما ألمحنا في نهاية القسم الفرعي ١,٣ من هذا الفصل فقد تزعم عالم الرياضيات الألماني النافذ «دِيفِيدْ هِيلْبِرْت» في القسم الثاني من عشرينيَّات القرن العشرين وبدايات ثلاثينيَّاته الدعوة إلى الإجابة على التساؤل المركزي «هل يمكن تأسيس منظومة الرياضيات بجميع فروعها وبأكملها على أساس من المنطق الرياضي الـمُحكَم من أجل ضمان اكتمالِها وخلوِّها من التناقض؟»، وجاءت الإجابة ما بين منتصف الثلاثينيات إلى منتصف الأربعينيات مفاجئةً ويمكن تلخيصُها في جُزْأين:

أ. بَرْهَنَ «كِيرْت غِيدِل» رياضيًّا على استحالة جمع أيةِ منظومةٍ رياضياتٍ لخاصِّـيَّتَيْ «الاكتمال» و «عدم التناقض» معًا، وبذلك بَيَّنَ أن هناك حدودًا لا يمكن للمنطق الرياضي أن يخترقها.

ب. في إطار هذه الحدود يمكن - من حيث المبدأ - مَيْكَنَةُ أي نَسَقٍ من المنطق الرياضي، وقد جاء ذلك نتيجةً لعمل «آلان تِيُورِنْغ» بنموذجه الحاسوبي النظري « آلة تيورنغ Turing Machine» و عمل «أَلُونْزُو تشيرش» بنموذجه البرامجي النظري «حساب لامْدَا Lambda Calculus».

وكان الجزء الثاني من الإجابة على سؤال «هِيلْبِرت» وقودًا مُشْعِلاً لحماس كبيرٍ بعد انتهاء الحرب العالمية الثانية؛ ففي النصف الثاني من أربعينيات القرن العشرين والنصف الأول من خمسينياته بدأ عدد متزايد من علماء الرياضيات، والهندسة الكهربية، والميكانيكية، والاقتصاد، وعلم النفس، والعلوم السياسية في نقاشات مكثفة حول إمكانية صناعة دماغ آلية. وكان من أمثلة فعاليات هذا النشاط ونتائجه ما يلي:

- بعد أن أظهرت أحدث مُكْتشَفات تشريح المنظومة العصبية في الحيوانات وفي البشر - في ذلك الوقت - أن المخ يتكون من شبكة كثيفة الترابط من الخلايا العصبية، دفعت هذه المكتشفات بعض الباحثين - مثل «مارْفِين مِينِسْكِي» أحد الآباء المؤسسين للذكاء الاصطناعي - إلى محاولة بناء شبكةٍ عصبيةٍ اصطناعيةٍ كهربائيةٍ عسى أن تظهر بعض قدرات المخ الطبيعي، واستطاعت هذه الشبكة العصبية بعد عمل مُطوَّل في تطويرها إظهارَ بعض التصرفات التي تكافئ العمليات المنطقية الثنائية البسيطة.

- على صعيد «المنطق الرمزي» ابتكر كلٌّ من «آلِنْ نِيُوِيلْ» و «هيرْبِرت سايْمون»[١] في عام ١٩٥٥م برنامجًا حاسوبيًّا سَمَّياه «الـمُنَظِّر الـمَنْطِقِيّ Logic Theorist» والذي استطاع في النهاية الوصول آليًّا بدءًا بالمسلَّمات الأولية إلى برهان ٣٨ من أصل ٥٢ «مُبَرْهَنَةً theorems» وَرَدَتْ في عمل «برتراند راسل» و «آلفريد نورث وايت هيد» – الذي أشرنا إليه في القسم الفرعي ١,٣ من هذا الفصل – وهو النجاح الذي جعل «هيربرت سايمون» يصرِّح بأن صناع هذا البرنامج قد توصلوا لحل اللغز القائم منذ القِدَم عن كيفية اكتساب آلةٍ ماديِّةٍ لعقلٍ مفكِّرٍ!
- في عام ١٩٥١م أنجز «كريستوفَر سترايتْشِي» في جامعة «مانشيستر» ببريطانيا برنامجًا حاسوبيًّا يستطيع أن يمارس لعبة «الضَّامَة/ الدَّامَا Checkers»، وفي السنة نفسها أنجز «دِيتْرِيشْ برِينْز» – الذي كان وقتها يتعاون مع جامعة «مانشيستر» أيضًا – برنامجًا يستطيع أن يمارس لعبة الشطرنج. وبالوصول إلى منتصف الخمسينيات من القرن العشرين كان الأمريكي «آرثر لِي صامْويل»[٢] قد أنجز برنامجًا يستطيع أن يتحدى لاعبًا هاويًا ذا مهارةٍ معتبَرةٍ في «الضَّامَة».[٣]
- اقترح «آلان تِيُورِنْغ»[٤] في عام ١٩٥١م اختبارًا للآلات الحاسوبية المفكِّرة، ويُعْرَف هذا الاختبار باسم «اختبار تيورِنغ Turing test»؛ حيث تخضَع أية آلة يزعم صانعوها أنها تمتلك تفكيرًا ذاتيًّا إلى عدد من المختبِرين (وهم أشخاص عادِيُّون) يدخل كُلٌّ منهم في حوارٍ نَصِّيٍّ عن بُعْدٍ مع هذه الآلة بتناوبٍ عشوائيٍّ مع محاوِرٍ بشريٍّ ويُطْلَبُ من المختبِر كلَّ فترةٍ أثناءَ الحوار أن يقرِّر ما إذا كان

١– وهو الحاصل على جائزة نوبل في الاقتصاد عام ١٩٧٨م.

٢– وهو من سَكَّ مصطلح «machine learning» وهو ما يمكن ترجمته إلى «التعلُّم التلقائيّ» أو «التعلُّم الحاسوبيّ» وقد اخترنا في هذا الفصل من الكتاب الترجمة الثانية، وسوف يلعب هذا المصطلح دورًا مُهمًّا فيما يلي من هذا الفصل.

٣– ومن وقتها صارت القدرات التنافسية للبرامج التي تمارس ألعاب الذكاء (مثل الشطرنج) أمام منافسيها من البشر مقياسًا لما وصل إليه مستوى التطور في مجال الذكاء الاصطناعي؛ فقد كان – على سبيل المثال – تغلُّب الحاسوب «Deep Blue» الذي طورته شركة «آي بي إم» في شهر مايو من عام ١٩٩٧م على الرُّوسيِّ الذي انفرد ببطولة العالم في الشطرنج عشرة أعوام متتالية «غاري كاسباروف» بـ «ثلاث جولات ونصف مقابل جولتين ونصف» علامةً مهمة على تطور حقل الذكاء الاصطناعي كله.

٤– وهو أحد الآباء المؤسسين لكُلٍّ من «علوم الحاسب» و «الذكاء الاصطناعي»، والذي ورد ذكره آنفًا أكثر من مرة في هذا الفصل.

محاورُه الآن بشريًّا أم آليًّا، وبناءً على التحليل الإحصائي لإجابات المختبِرين تُعْطَى كُلُّ آلةٍ درجةً لتقييم ذكائها. ورغم الانتقادات اللاحقة التي وُجِّهَتْ إليه، فلا يزال يُعَدُّ «اختبارُ تيورنج» أولَ طَرْح جاد في «فلسفة الذكاء الاصطناعي».

مع تزايُدِ هذا النشاط البحثي وتراكم إنجازاته كان من المهم أن تتداعى الشخصيَّات العلمية البارزة وَقْتَها في هذا الخضم إلى تدارس هذا الحقل الجديد الواعد تدارُسًا مُطوَّلاً معمَّقًا، فكان أَنْ عُقِدَ مؤتمَر «دارْتمُوث»[1] في صيف ١٩٥٦م الذي يُعَدُّ بمثابةِ الميلاد الأكاديمي للذكاء الاصطناعي حيث سُكَّ الـمُصْطَلَح.

وفي الحقبة التي تلت هذا المؤتمر ارتفعت موجة البحث والتطوير مدعومةً بمِنَحٍ ماليةٍ من «وكالة مشروعات الأبحاث الدفاعية المتقدمة (دارْبَا) DARPA» في الولايات المتحدة الأمريكية لعدد من كبرى الجامعات الأمريكية؛ مثل «معهد ماساتشوسيتس للتقنية (إم آي تي) MIT» وكذلك «جامعة ستانْفورد» و «جامعة كارْنيغِي مِيلُون» مما مكنها من إنشاء برامج أو أقسام أو حتى معاهد مكرسة لتخصص «الذكاء الاصطناعي»،

١- هو ورشة عمل طويلة استمرت لمدة شهر ونصف في جامعة «كُلِّيَّة دارْتمُوث» في مدينة «هانوفَر» بولاية «نيوهامْشِير» بالولايات المتحدة الأمريكية، وفيه اجتمع أحد عشر من علماء الرياضيات وعلوم الحاسب للقيام بالنقاش والعصف الذهني المطوَّل. وقد صار نَصُّ المقترَح الأصلي لورشة العمل وثيقةً تاريخيةً يستحق أن نُورِدَه فيما يلي لفائدة موضوع هذا الفصل وقارئه:

“We propose that a 2-month, 10-man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer”.

وترجمتُهُ للعربية هي:

«نقترح العكوف على دراسةٍ للذكاء الاصطناعي تمتد لشهرين بعشرة مشاركين تُعْقَدُ في صيف ١٩٥٦م في كلية دارْتمُوث في مدينة هانوفَر بولاية نِيُوهامْشِير. وسوف تسير الدراسة على أساسِ الافتراض القائل بأن أي مَلْمَحٍ من ملامح عملية التعلُّم أو أي مَلْمَحٍ آخر للذكاء يمكن من حيث المبدأ وَصْفُهُ بدقةٍ تكفي لجعل الحاسوب قادرًا على محاكاته. كما سنقوم بمحاولةٍ لاستكشاف كيفية جعل الحواسيب قادرةً على استعمال اللغة، وتكوين المفاهيم المجرَّدة، وحل المسائل التي لا يقدر على معالجتها الآن إلا البشر، وتطوير الحواسيب لذواتها بذواتها. ونظن أنه يمكن إنجاز تقدم معتَبَر في واحدةٍ أو أكثر من هذه المسائل إذا عكفت مجموعة من العلماء المنتقَيْن بعناية على العمل عليها معًا طِيلةَ الصيف»

كما يُذْكَر «مختبَر الذكاء الاصطناعي» بجامعة «أَدِنْبِره» في «اسكتلندا» في ذلك الوقت كمثال على مراكز الذكاء الاصطناعي الهامة خارج الولايات المتحدة الأمريكية.[١]

ولذلك يطلق العديد من مؤرِّخي العلوم اسم «السنوات الذهبية» للذكاء الاصطناعي على تلك الحقبة الممتدة من عام ١٩٥٦م إلى عام ١٩٧٤م، وفيما يلي سوف نحاول وضع القارئ في بعض أجواء ما ساد في تلك الفترة الحافلة من أساليب معالجة[٢] المسائل المطروحة وبعض الإنجازات التي تحققت أثناء تلك «السنوات الذهبية»:

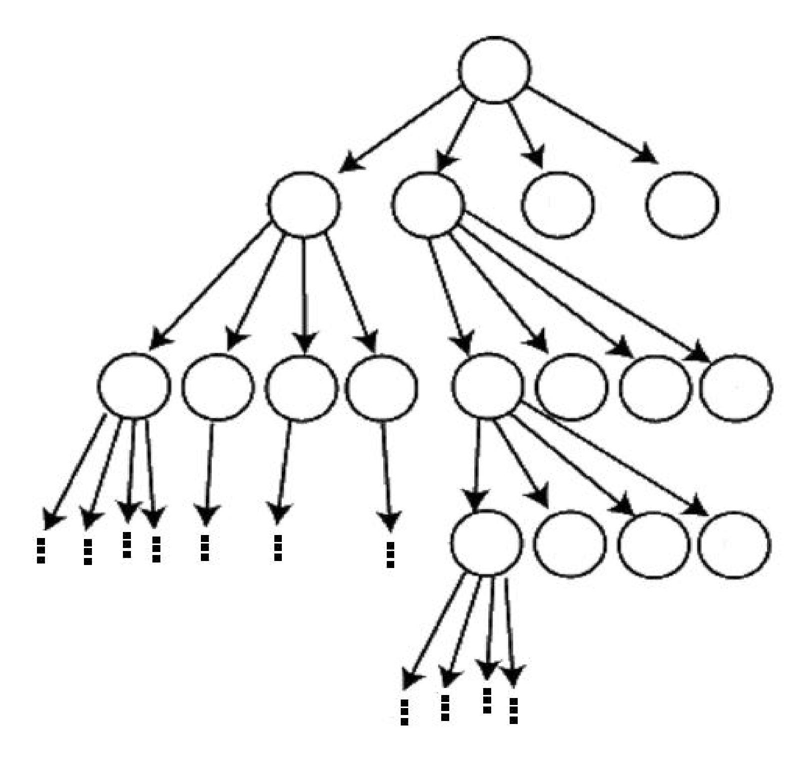
- صياغة ومعالَجة المسائل كعمليةِ بَحْثٍ شَجَريٍّ: حيث يعتمد هذا الأسلوب على رُكْنَيْن أولهما هو صياغة المسألة محل الدراسة على هيئة «حالات states» بدءًا بـ «حالةٍ ابتدائيةٍ initial state» مصحوبةً بآليةٍ لتوليد الحالات التي يمكن أن تقود إليها كل حالة ومع تعريف الحالات التي تُعَدُّ حَلًّا مقبولاً للمسألة. أما الركن الثاني فهو إجراءُ عملية «البحث الشجري tree search» التي تُفْتَتَحُ بالحالة الابتدائية للمسألة والتي إن لم تمثل حلًّا مقبولاً يتم توليد الحالات التي تقود إليها الحالة الابتدائية، ومن ثم يجري اختبارُ كل حالةٍ مولَّدةٍ من حيث كونها حلًّا مقبولاً أم لا، ثم يتوالى هذا التوليد والاختبار للحالات جَوْلةً وراءَ أخرى حتى يتم الحصول على أحد الحلول المقبولة (أو حتى تحصيلِ عددٍ معرَّفٍ ابتداءً من تلك الحلول) أو حتى يتم توليد عددٍ تراكميٍّ أقصى معرَّفٍ ابتداءً من الحالات وحينَها يتم الإعلان عن عدم التوصل إلى حل للمسألة[٣].

ويصور (الشكل ٢) المجاور مثالاً لشجرة الحالات الناتجة عن عملية البحث الشجري؛ حيث تمثَّل كُلُّ حالةٍ على هيئة كرةٍ صغيرةٍ، وتُرْسَمُ الحالة الابتدائية (جذر الشجرة) عادةً في أعلى الشجرة، وتُمَثَّلُ العلاقة بين كل حالة والحالات التي يمكن أن تتولد منها بأسهم في اتجاه هذا التولُّد، وتُسَمَّى الحالات التي لا تخرج منها أية أسهم «حالاتٍ طَرْفيَّة terminal states».

١- بالتوازي مع ذلك كان اليابانيون في تلك الحقبة وهم يعيدون بناء الاقتصاد والصناعة في بلادهم بعد دمار الحرب العالمية الثانية يركزون جهودَهم على تطوير الروبوتات الصناعية.

٢- والتي لا بُدَّ للقارئ أن يُلِمَّ بها من أجل استيعاب المواضيع التي تستعرضها الأقسام التالية من هذا الفصل - وخصوصًا قسمُهُ الثاني - وقد اجتهدنا قدر الإمكان من أجل عرضها بوضوحٍ وتبسيطٍ غير مُخِلٍّ للقارئ غير المتخصص.

٣- يتم أيضًا الإعلان عن عدم التوصل إلى حلٍ للمسألة عند عدم تولُّد أية حالات جديدة.



الشكل ٢: مثال على «شجرة البحث Search tree».

وقد تفنن باحثو الذكاء الاصطناعي وتوسعوا في ابتكار خُوارِزمات البحث الشجري بهدف التوصل إلى حل/ حلول مقبولة (إن وُجِدَت) للمسألة بأسرع ما يمكن عَبْرَ الاقتصاد في توليد حالات المسألة قَدْرَ الإمكان بتوجيهِه أولاً ناحية المسارات التي يُتَوَسَّمُ فيها تحقيقُ أقصى اقترابٍ من الحل[1]؛ فتجاوزوا أساليب البحث العمياء مثل «البحث الشامل exhaustive search» وابتُكِرَت خُوارِزماتٌ بارِعةٌ؛ مثل «مُونْت كارْلُو Monte Carlo» وكذلك «A* search».

كذلك ابتُكِرَت خُوارِزمات بحث شجري لاختيار أفضل اختيارٍ/ نَقْلةٍ خاصةٍ بالألعابِ التي يتنافس فيها الحاسوب مع خصم آخَر - سواء كان بشريًّا أو برنامجًا جاسوبيًّا آخَر - بالتناوب (مثل الشطرنج) ومن أبرزها خوارزم «ألْفا-بِيتَا Alpha-Beta». والواقع أنه مع بدء سبعينيات القرن العشرين الميلادي صار

١- وهو ما يُعْرَفُ إجمالاً بـ«heuristic search» في مقابِل البحث غير الموجَّه (الأعمى) الذي يُعْرَف بـ«blind search».

البحث الشجري وخوارزماته رُكْنًا أساسيًّا كبيرًا من مباحث الذكاء الاصطناعي مؤسَّسًا على صياغة رياضية مُحْكَمة.

وبالاعتماد على خُوَارِزْمات البحث الشجري هذه طُوِّرَتْ العديدُ من البرمجيات ذات القدرات الـمُبْهِرةِ «آنَذاكَ»؛ فاستطاع بعضُها أن يواجِهَ - ولو بأداءٍ أوَّليٍّ - خصمًا في الألعاب التناوبية، كما تمكنت أخرى من حَلِّ بعض المسائل في الجبر والهندسة الإقليدية[١]، وأخرى تستطيع إجراءَ «التكامُلات بالرموز symbolic integrals»[٢]، وأخرى لتخطيط حركة الروبوتات[٣].

- ميكنة المنطق الرياضي والتوسع في حوسبته: بينما شهدت العقودُ الأخيرةُ من القرن التاسع عشر والعقود الأولى من القرن العشرين الميلادي تنظيرَ المنطق تنظيرًا رياضيًّا عميقًا فاستُكْشِفَت حدودُه وصار من الفروع النظامية الأساسية للرياضيات - كما أشرنا في القسم ١,٣ آنفًا - فإن «السنوات الذهبية» قد شهدت حوسبة هذا المنطق الرياضي[٤] بتوسُّعٍ ليصير أحد الأدوات الرئيسية التي يرتكز عليها حقل «الذكاء الاصطناعي» ويوظفها بكثافةٍ في تطبيقاتِه التي دَعَّمَت الثقة في القدرات الواعدة لهذا الحقل الناشئ؛ حيث ظهرت برمجيات تمكنت ابتداءً من «المسلَّمات الأولى axioms» من الوصولِ إلى إثبات بعض «مُبَرْهَنات theorems" الهندسة الإقليدية والجَبْر الأوَّلِيِّ وحَلِّ بعض مسائلِها

١- ومن أمثلتها «Geometry Theorem Prover» من تطوير «هيربِرْت غِيلِيرْنَرْ» أستاذ «علوم الحاسب» بجامعة «ستوني بروك» في «نيويورك» بالولايات المتحدة الأمريكية عام ١٩٥٨م.

٢- ويعرف أحد أمثلتها الباكرة باسم «Symbolic Automatic INTegrator (SAINT)».

٣- ومن تلك البرمجيات «Stanford Research Institute Problem Solver (STRIPS)» الذي جرى تطويره في جامعة «ستانفورد» الشهيرة في الولايات المتحدة الأمريكية عام ١٩٦١م والذي جرى توظيفه كاداةٍ للتحكم في سلوك الروبوت «شاكِي Shakey» الذي كانت قد طورته الجامعة وقتذاك.

٤- المنطق الرياضي المقصود هنا لا يشمل فقط «المنطق الأَرِسْطِيّ» الذي من أمثلته: المقدِّمتان <<كل إنسان فانٍ>> و <<سالم إنسانٌ>> فتَتْبَع ذلك النتيجة <<سالمٌ فانٍ>>، فهذا المنطق المعروف أيضًا باسم «المنطق اللفظي/ الكلامي propositional logic» يُعَدُّ في منظومة المنطق الرياضي الآن بمثابة «الدرجة الصِّفرية zeroth order» وهو ذو قدرة محدودة في التعريف والاستنتاج. أما االمنطق الرياضي من «الدرجة الأولى first order» والمعروف أيضًا باسم «Predicate Calculus» فهو أقوى كثيرًا في قدرته على التعريف والاستنتاج والتعميم بسبب اشتماله على دوالَّ ومتغيراتٍ وهو الأكثر شيوعًا في البرمجة المنطقية الرياضية الآن، كما أن هناك المنطق الرياضي من درجاتٍ أعلى (ثانيةٍ فما فوق) الذي يسمح بأن تكون بارامِتْراتُ دَوَالِّه دَوَالًّا في حد ذاتها.

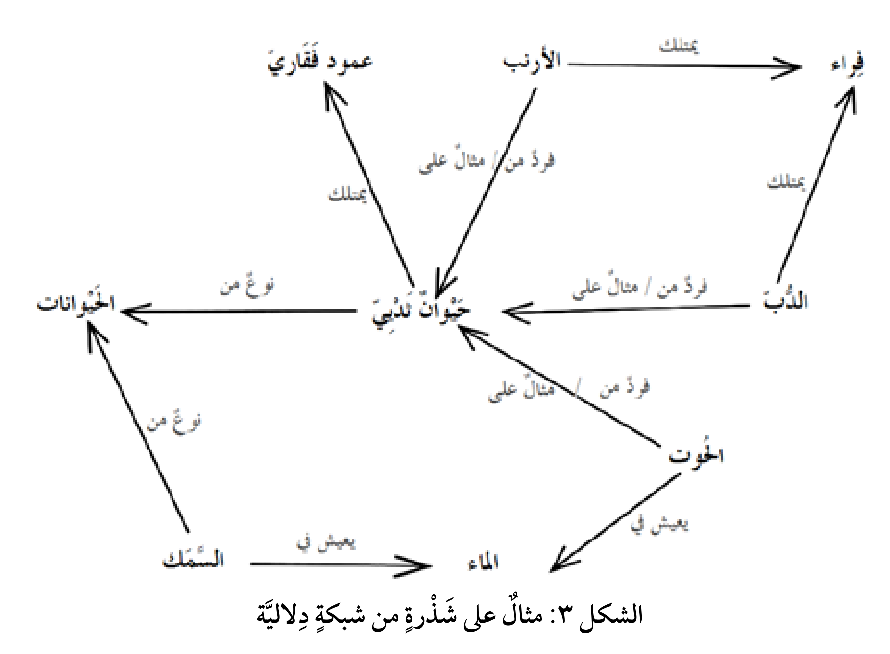
كذلك بالتزاوج مع خوارزمات البحث الشجري كما أشرنا في الفقرة السابقة، وتُوِّجَ ذلك مع نهايات تلك الحقبة – وتحديدًا عام ١٩٧٢م – بميلاد لغة برمجة الحواسيب المعروفة باسم «Prolog» – اختصارًا لعبارة «البَرْمَجَة عبر المنطق Logic Programming» – التي تُيَسِّر لمطوري البرمجيات استدعاء وتوظيف أدوات المنطق الرياضي في برمجياتهم بسلاسةٍ.

- المحاولات الأولى لمعالجة اللغات الطبيعية: فيما كان طبيعيًّا أثناء تلك السنوات الذهبية للذكاء الاصطناعي أن تكون ميكنةُ تحليلِ وتوليدِ «اللغات الطبيعية natural languages» في بؤرة اهتمام الباحثين، فقد كان لافتًا وقتها ذلك النضج المبكر المتمثل في تجاوُز المستويات السطحية الشَّكْليَّة لِلُّغةِ الطبيعية والولوج مباشرةً إلى تناول مستوياتها العميقة وتحديدًا المستوى «الدِّلالي semantic» و«الأُنْطولوجي ontological».

فلَمَّا كانت «نظرية الاتصال المعلوماتي information theory» التي تأسست واستقرت قبل تلك الحقبة بعِقْدَين وحققت نجاحات هائلة في هندسة الاتصالات وعلوم الحاسب قد عالجت مسائل الترميز على مستوًى متقدمٍ وطورت في ذلك السياق آلياتٍ رياضيةً رفيعةً، فإن الباحثين نظروا إلى المستويات الشكليةِ من اللغة الطبيعية (مثل «النَّسَق الكِّتابيّ orthography»، و «النَّسَق الصَّرْفيّ morphology») على أنها مسائل ترميزية تسهل معالجتها بنفس آلياتِ نظرية الاتصال المعلوماتي. وبالتوازي مع ذلك كان العمل في ستينيات ذلك القرن جاريًّا على قدمٍ وساقٍ في تطوير لُغَاتِ بَرْمَجَةٍ للحواسيب الرقمية تشابه إلى حَدٍّ كبيرٍ لُغَةَ توصيف الخوارزمات الرياضية لتيسير عملية تطوير البرمجيات وإبعادها عن تفصيلات «العتاد hardware» الإلكتروني الذي تتكون منه الوحدات الحِسابية المنطقية ووحدات التحكم ووحدات الذاكرة في مختلَف تلك الحواسيب.[1] وكان من لوازم ذلك العمل تطويرُ تقعيداتٍ

١– وذلك من أجل توسيع قاعدة المبرمِجِين، ورفع إنتاجية عملية تطوير البرمجيات، وتقليل الأخطاء فيها، وكذلك – وربما كان هذا هو الأهم اقتصاديًّا – جعل نفس البرمجيات قابلة للاستخدام على حواسيبَ رقميةٍ مختلفةٍ في معمارِ عَتَادِها الإلكتروني وهي الخاصية المعروفة باسم «portability». وسنتوقف مجدَّدًا مع لغات البرمجة هذه في القسمين الثاني والثالث من هذا الفصل.

وأدواتٍ رياضيةٍ رفيعةٍ لمعالَجَةِ العديد من الجوانب المعجَمية والنحوية من اللغة – إضافةً إلى الجوانب الترميزية التي أشرنا إليها أعلاه – وهو ما افترض باحثو الذكاء الاصطناعي – بتفاؤلٍ مُفْرِطٍ – وقتذاك أنها قابلةٌ أيضًا للتطبيق على اللغات الطبيعية.



الشكل ٣: مثالٌ على شَذْرةٍ من شبكةٍ دِلاليَّة

ولذلك انصبت أبحاثُ الذكاء الاصطناعي في معالَجَة اللغات الطبيعية على مستواها الدِّلاليِّ حيث تُجَمَّع الوحداتُ الترميزية لِلُّغَة («مفرداتُها» على سبيل المثال) تحت «مفاهيم concepts» ثم تُدْرَس وتوصَّف العلاقاتُ بين هذه المفاهيم بروابطَ دِلاليةٍ فيكون حاصلها ما استقرت تسميته بـ «الشبكة الدِّلالية semantic network»، ويصوِّر (الشكل ٣) أعلاه شذرةً صغيرةً مبسَّطةً من مثل تلك الشبكة الدلالية. ويُنْسَبُ ابتكار الشبكة الدلالية المحوسَبة إلى «ريتشارد رِيتْشِنز» عام ١٩٥٦م من جامعة «كامْبِريدْج» الإنغليزية لتكون جسرًا بين اللغات المختلفة في الترجمة الآلية.

أما الدرس الأونطولوجي فيُعْنَى بوصف العالَم الواقعي حاسوبيًّا وربطِ وحدات اللُّغة بهذا النموذج الحاسوبي للعالم، وقد كان ذلك – ولا يزال – تحدِّيًا كبيرًا حتى لو

لم يكن هذا العالم الواقعي المطلوب وصفه حاسوبيًّا هو العالم الحقيقي بأكمله بل فئةً جزئيةً محدودةً منه.[١]

ومن أمثلة التطبيقات التي أظهرت وَقْتَها تعامُلاً لافتًا مع اللغة الطبيعية برنامجٌ حاسوبيٌّ سُمِّيَ «الطالب STUDENT» طوره «جورج بُوبْرُو» ضمن أطروحته للدكتوراة في «معهد ماساتشوسيتس للتقنية MIT» عام ١٩٦٤م بقدرته على حل بعض المسائل الكلامية في الجبر من مستوى المدارس الثانوية، وبرنامجٌ حاسوبيٌّ طَريفٌ آخرُ سُمِّيَ «إلِيزا ELIZA» طُوِّرَ في الفترة بين ١٩٦٤م و ١٩٦٦م أيضًا في «معهد ماساتشوسيتس للتقنية» يستطيع إجراءَ محاوَراتٍ «نصيةً» مع مستخدِمين بَشَريِّين ظَنَّ أغلبُهم أن محاورهم بشريٌّ وليس آلةً حيث بدا أداء «إليزا» طبيعيًّا للغاية وإن كان ذلك نتيجةَ استخدام بعض الحيل البسيطة كالأجوبةِ المعلَّبة العامَّةِ وكذلك تكرار أسئلة المستخدِمين في الردود عليهم.[٢]

- التوجُّه لدراسة الذكاء الاصطناعي في ظُروفٍ مَعْمَليَّةٍ مقيَّدةٍ: وهو الـمَلْمَحُ الأخيرُ الذي ننوه إليه هنا من تلك الملامح التي صبغت أجواء تلك «السنوات الذهبية» للذكاء الاصطناعي؛ حيث نادت في نهاية الستينيات من القرن العشرين الميلادي أصواتٌ مسموعةٌ في حقل الذكاء الاصطناعي من أبرزها «مارْفِين مِينِسْكِي»[٣] و«سايْمُور بابِيْرت» وكلاهما من «معهد ماساتشوسيتس للتقنية MIT» إلى الانكباب على تطوير خُوارِزْمات الذكاء الاصطناعي - وهو ما زال حقلاً ناشئًا - على مسائلَ كائنةٍ في بيئةٍ مثاليَّةٍ (افتراضيةٍ)[٤] إلى أن تُفْهَمَ

١- هناك بالطبع مستوًى آخر للغة وهو «اللغة المنطوقة speech» استماعًا ونُطْقًا، وقد اتخذت معالجة اللغات الطبيعية وقتذاك على هذا المستوى - لحسن الحظ - مسارًا مختلفًا بعيدًا عن الذكاء الاصطناعي، وقد لعب ذلك المسار دورًا مهمًّا فيما بعد، سوف نأتي على ذكره في الأقسام التالية من هذا الفصل وخصوصًا من الثاني للرابع.

٢- وهذا مثالٌ على النقدِ الموجَّهِ لاختبار «تِيُورِينْغ» لتعيين درجةِ ذكاءِ الآلاتِ المفكِّرةِ المشار إليه آنفًا في هذا القسم الفرعي.

٣- وهو أحد الاثنين الذين دَعَوَا إلى مؤتمر «دارْتمُوثْ» الذي أشرنا إليه آنفًا في هذا القسم الفرعي، وهو كذلك أحد مؤسِّسي مُخْتَبَر الذكاء الاصطناعي في «معهد ماساتشوسيتس للتقنية MIT»، ويُنْظَر إليه الآن على أنه أحد الآباء المؤسسين لهذا الحقل.

٤- وكانت الحجة القوية لأصحاب هذه الدعوة أن هذه المقاربة قد اتُّبِعَت بنجاح في العلوم الأخرى؛ فالفيزياء النظرية على سبيل المثال قد استخدَمت مفاهيمَ مثاليةً كالغازات المثالية، والسوائل المثاليةً، والسطوح المثالية، والجسم الأسود، ... إلخ لتطوير قوانينها الأساسية، ثم استُخدِمَتْ هذه القوانينُ فيما بعد بنجاحٍ عريضٍ في ما يصعب حصره من التطبيقات الهندسية بعد الأخذ في الحُسْبان تأثيرَ الفوارق بين الافتراض المثالي والواقعِ الحقيقي.

ديناميكياتها بعمق وإلى أن تنضج كفاءة وموثوقية هذه الخوارزمات وحينئذٍ يُمكِنُ تطبيقُها على تلك المسائل في بيئاتها الحقيقية بتفاصيلها المعقَّدة. وكانت البيئة المثالية الـمُجرَّدة المقترَحة هي كُتَلٌ من أشكالٍ وأحجامٍ وألوانٍ مختلفةٍ موزَّعةً على سطح مستوٍ، وبالفعل انصب الكثير من النشاط البحْثي على معالجة المسائل في إطار هذه البيئة المثالية. واستطاع «مينسكي» و «بابيرت» الداعيان لهذه المقاربة تطبيقها لبناء ذراع رُوبُوتيةٍ تمكنت بنجاح من التقاط ورص هذه الكتل في تشكيلات معقَّدةٍ نسبيًّا.

وربما كان النجاح الذي توج كل ما سردناه آنفًا من ملامح تلك المرحلة هو البرنامج الحاسوبي «SHRDLU» الذي طوره بين عامَيْ ١٩٦٨م و ١٩٧٠م «تيري آلين وِينوغْراد» أستاذ علوم الحاسب في «جامعة ستانفورد» الأمريكية والذي كان يستطيع التواصل بالإنغليزية الاعتيادية مع مُستخدِميه ويستقبل طلباتهم بأداء مهامَّ معيَّنةٍ لمناوَلة الكُتَل في البيئة المثالية الموصوفة آنفًا، ومن ثم يقوم بتخطيط خطوات تحقيق المناوَلة المطلوبة وتنفيذِها على شاشة رسومية.

وفي خِضَمِّ تلك الحقبة كان التفاؤلُ بمستقبل الذكاء الاصطناعي مفرِطًا ويكفي ذِكْرُ بضع اقتباسات من تصريحات كبار الباحثين وقتذاك كي يلمس القارئ مدى الاندفاع في ذلك التفاؤل:

- ذكر «هيرْبرت سايمون ألِكْساندَر» [١] و «آلين نِيُوِيل» [٢] في ورقةٍ نُشَرت لهما في دورية «بحوث العمليات Operations Research» عام ١٩٥٨م أنه «في غضون عشر سنواتٍ سوف يكون بطل العالم في الشطرنج حاسوبًا رقميًّا، وفي غضون عشر سنواتٍ أيضًا سوف يكتشف حاسوبٌ رقميٌّ مُبَرْهَنَةً رياضيةً رئيسيةً جديدةً ويُثْبِتُ صحتَها».
- أكد «هيرْبرت سايمون ألِكْساندَر» في مؤلَّفه «طبيعة مَيْكَنة الأعمال من أجل البشر والإدارة» المنشور عام ١٩٦٥م أنه «في غضون عِقْدين من الزمان سوف تصير الآلات قادرةً على أداء أي عمل يستطيع الإنسان أداءه».

١- وكان أستاذًا بجامعة «كارْنيغي مِيلون» الأمريكية، وهو أحد الباحثين الرواد والمفكرين النافذين في حقل الذكاء الاصطناعي في حقبته الأولى، وهو أيضًا الحاصل على جائزة نوبل في الاقتصاد عام ١٩٧٨م، وقد سبق ذكره في هذا الفصل.

٢- وكان أستاذًا في «علوم الحاسب» و«علم النفس الإدراكي» بجامعة «كارْنيغي مِيلون»، وقد سبق ذكره في هذا الفصل.

- أكد كذلك «مارفين مينسكي» في مؤلَّفه «الحوسبة: آلاتٌ محدودةٌ ولا نهائية» المنشور عام ١٩٦٧م أنه «خلال جيلٍ واحدٍ فقط ... فإن مسألة تخليق ذكاءٍ اصطناعيٍّ سوف تكون قد حُلَّتْ إلى حَدٍّ بعيد».
- وفي مقابلةٍ له عام ١٩٧٠م مع «مجلة الحياة Life Magazine» الأمريكية الشعبية النافذة وقتذاك صرح «مارفين مينسكي» أنه «في غضون من ثلاث إلى ثماني سنوات فإننا سنتمكن من بناء آلةٍ ذات مستوى ذكاءٍ عامٍّ مساوٍ لذكاء الإنسان العادي».

فهل صَدَّقَت الأحداث التالية ذلك التفاؤل الكبير؟

٧,١. السنواتُ العِجافُ، ونَقْدُ السنوات الذهبية.

لا رَيْبَ أن القارئ الذي صحبنا حتى هذه النقطة من هذا الفصل قد لاحظ أن مَهْدَ الذكاء الاصطناعي ومَرْكَزَ الثقل الأكبر لأنشطة البحث والتطوير فيه من خمسينيات إلى سبعينيات القرن العشرين الميلادي هو الولايات المتحدة الأمريكية[١] حيث تحصل الأبحاث الجامعية على غالب تمويلها من الـمُجَمَّع الصناعي العسكري. ولذلك فإن استمرار تمويل هذه الأبحاث مشروطٌ بتحقيق هذه الأبحاث لمستهدَفات تطوير التقنيات الصناعية و/ أو العسكرية التي وَعَدَتْ بها وَقْتَ حصولها على تلك التمويلات؛ فعلى سبيل المثال كانت «وكالة مشروعات الأبحاث الدفاعية المتقدمة (دارْبَا) DARPA» الأمريكية - وهي إحدى كُبْرَيات الجهات الممولة لهذه الأبحاث - تحت ضغط السياسيين كي تَقْصُرَ تمويلاتها على المشروعات ذات المهامِّ العَمَلِيَّةِ المحدَّدةِ بوضوحٍ والتي تلتزم بتنفيذها بدِقَّةٍ لا سيما بعد تمرير التعديل القانوني في هذا الاتجاه المنسوب للسيناتور الأمريكي «مانِسْفِيلْد» سنة ١٩٦٩م، وبطبيعة الحال فإن أكبر الضحايا جَرَّاءَ ذلك هي الأبحاث الاستكشافية المفتوحة التي تُعَدُّ وفق هذا التَّوَجُّه أنشطةً «نظريةً بَحْتة»!

١ - تليها بمسافةٍ غير قصيرةٍ حليفتُها العسكرية والاقتصادية الأولى في تلك الفترة «المملكةُ المتَّحدة».

وفي حين أخذت «السنوات الذهبية» [1] تقترب من نهايتها، فإن أجواءً متصاعدةً من الإحباطِ بل السخطِ كانت تتصاعدُ في أوساط الجهات الممولة لأبحاث الذكاء الاصطناعي[2]، وبالوصول إلى عام ١٩٧٤م انتهت تلك الجهات المموِّلة إلى قناعةٍ بأن حقل الذكاء الاصطناعي قد فشل فشلاً ذريعًا في تحقيق الآمال العريضة المعلَّقة عليه قياسًا على التوقعات ذات السقوف العالية التي كانت تروِّجها بثقة كبارُ الشخصيات المتصدرة للمشهد البحثي - مثل تلك التي وردت في ختام القسم الفرعي السابق. ومنذ ذلك الوقت ولِعِدَّةِ سنواتٍ لاحقةٍ توقفت المِنَحُ التمويليةُ لأية مقترحاتٍ بحثيةٍ في إطار «الذكاء الاصطناعي» توقفًا شبه كامل استمر حتى سنة ١٩٨٠م.

فبالرغم من الإنجازات العديدة المبهِرة التي حققها هذا الحقل الناشئ في تلك الحقبة - مما أشرنا إليه في القسم الفرعي السابق - والتي روجت لقدرته في المستقبل على صنع المعجزات، فإن إبهارها ذلك قد نبع من مقارنتها بلا شَيْءَ قَبْلَها، أما بالمقارنة مع التوقعات الـمُسرِفة في التفاؤل التي كانت تُطْلَقُ آنَذاكَ بالوصول إلى حلولٍ ناجعةٍ في وقت قريبٍ للمسائل الكبرى المتصلة بمحاكاة الذكاء البشري فإن تلك «الإنجازات» تُعَدُّ بمثابة «ألعاب الأطفال» على حد وصف تقاريرَ رسميةٍ صدرت في نهاية ستينيات وبداية سبعينيات القرن العشرين الميلادي لرفع واقع أبحاث الذكاء الاصطناعي على وجه العموم أو بعض مسائله على وجه الخصوص؛ ومن أمثلتها المؤثرة: «تقرير أَلْباكْ (Automatic Language Processing Advisory Committee (ALPAC» الذي أصدرته لجنة علمية مكلَّفة من قبل الحكومة الأمريكية في النصف الثاني من الستينيات لتقييم التقدم الذي حققته «اللسانيات الحاسوبية» عمومًا و«الترجمة الآلية» خصوصًا، وتقرير «لايتْ هِيلْ Lighthill» عام ١٩٧٣م لتقييم كامل أوضاع أبحاث الذكاء الاصطناعي في بريطانيا.

لم تقتصر تلك الانتقادات القاسية وقتئذٍ على التقارير الرسمية بل إن الساحة العلمية والفكرية حفلت بالعديد من الفلاسفة والكثير من كبار المتخصصين في الرياضيات وعلوم الحاسب والاتصالات وعلم النفس الإدراكي ... - بل وحتى باحثين في الذكاء الاصطناعي

١- وهي الفترة بين أعوام ١٩٥٦م و ١٩٧٤م كما أسلفنا في القسم الفرعي السابق.

٢- ومن أهمها في تلك الفترة بالإضافة إلى «دارْبا» الأمريكية: «مجلس الأبحاث الوطني(National Research Council (NRC» في الولايات المتحدة الأمريكية، و«مجلس الوزراء البريطاني» في المملكة المتحدة.

نفسِهِ – ممن تصدَّوْا لنقد آليات الذكاء الاصطناعي والتفنيد في ضوء هذا النقد لمزاعم بعض نجوم الذكاء الاصطناعي بقدرته خلال وقت قريب على محاكاة الذكاء البشري.

وبينما أدت تلك الحملة المكثَّفة التي سُطِّرَت فيها أكوامٌ من المقالات والأوراق البحثية والكتب إلى انتقاصٍ كبيرٍ من مصداقيةِ حقل الذكاء الاصطناعي استمر قائمًا لعدة سنوات تالية، فإن ما أنتجته من تشخيصٍ تفصيليٍّ لمشاكلِ الأساليب المتَّبَعة وقتئذٍ في هذا الحقل – والذي استقر من وقتها في أدبيات علوم الحاسب عمومًا والذكاء الاصطناعي خصوصًا – كان مَعْلَمًا فارقاً في ترشيد مسيرة هذا الحقل وتطويره كثيرًا فيما بعد؛ بل استمر أثره حتى وقتنا الحاضر. وفيما يلي نوجز أهم تلك العقبات التي كان على الذكاء الاصطناعي أن يضعها بعين الاعتبار ومن ثم يحاول تجاوزها حتى يواصل مسيرته نحو تحقيق إنجازاتٍ صناعيةٍ وخَدَميةٍ ملموسةٍ ذات آثارٍ اقتصاديةٍ ومَعْرِفيةٍ ملموسة في العالم الواقعيِّ:[١]

- كان «**البحثُ الشَّجَريُّ**» السلاحَ الأساسيَّ الضاربَ الذي وظَّفَه الباحثون آنذاك لتوصيف وحل الكثير من – أو بالأصح «أغلب» – المسائل التي تصدَّوْا لها، وفي حين أن هذا الأسلوب يتمتع بميزات: «البساطةِ كنموذجٍ رياضيٍّ حاسوبيٍّ»، و «سهولة التنفيذ برامجيًّا»، و «العُمومية» [٢]، ولكنه في الوقت ذاته يعاني من عيوبٍ ثلاثةٍ (على الأقل) يطعن كلٌّ منها بشدةٍ في فعالِيَتِه:

الأول: التضخُّمُ الأُسِّيُّ للعدد الذي يجري توليده من حالات المسألة في شجرة البحث؛ فإذا رُمِزَ لمتوسط عدد [٣] الحالات التي تولِّدها كل حالةٍ في شجرة البـــحث بالرمـــز م وإلى العمق المطلوب استكشافه حتى الانتهاء من المسألة

١- وتخص جميعُ هذه العقبات خُصوصيَّةً وثيقةً المعالجةَ الحاسوبيةَ لِلُّغاتِ الطبيعيةَ، ولا مفر من فهمها بعمق ومن ثم تفاديها أو ترويضها لمن أراد أن يخوض بجديةٍ مضمار هذه المعالجة من باب البحث والتطوير. ولذلك نُبْرِزُ هذه العقباتِ عَبْرَ نَقْدٍ موجَزٍ لبعض الآليات الأساسية التي ارتكز عليها هذا الحقل في تلك الحقبة، وقد قَدَّمْنا لبعض هذه الآليات بإيجازٍ في القسم الفرعي السابق من هذا الفصل، وسوف نعاود في هذا الفصل ولا سيما قسمَيْه الثاني والثالث التعرضَ لتلك العقبات والتعامل معها.

٢- بمعنى أن خُوارِزْمَ البحثِ الشجريِّ ذاتَهُ يعاد استخدامُه مع مسائل مختلفة ويجري فقط تغييرُ «الإجراء routine» المسؤول عن توليد حالاتها، وكذلك الإجراء المسؤول عن اختبار كل حالةٍ من حيث كونِها حلًّا مقبولاً للمسألة أم لا.

٣- المقصود هنا ليس «المتوسط الحسابي arithmetic mean» ولكن المقصود هو «المتوسط الهندسي geometric mean» هـ الذي نحصل عليه لعددين $د_{١}$ ، $د_{٢}$ وفق الصيغة الحسابية: $هـ = (د_{١} \times د_{٢})^{(١ \div ٢)}$ ، وعلى نحو العموم فإن المتوسط الهندسي لمجموعة أعداد $\{د_{١}، د_{٢}، د_{٣} ...، د_{ن}\}$ يُحْسَب وفق الصيغة: $هـ = (د_{١} \times د_{٢} \times د_{٣} \times ... \times د_{ن})^{(١ \div ن)}$.

بالرمز ع فإن إجمالي عدد الحالات المطـلـوب توليدها واختبارها أثناء البحث = (مع – ١) ÷ (م – ١) [١] فإذا لم تكن المسألة «ضحلة» العمق، فإن أي حاسوب ستتقاصر قدراتُه سريعًا عن التوغل في عمق هذا البحث مهما عَظُمَتْ تلك القدرات الحاسوبية.[٢]

الثاني: في أغلب المسائل غير التافهة المُنَمْذَجةِ في صيغةِ البحث الشجري فإن حجم وعمق شجرة البحث حتى العثور على حلٍّ للمسألة (أو استيفاء حلولها) ليس معلومًا قبل إجراء عملية البحث، ولذلك فإن التوقف عن البحث عند حد معين دون العثور على حلول بسبب نفاد الزمن المحدد أو الموارد الحاسوبية يُنْتِجُ إجابةً «غير يقينية undecided» بين وجود أو عدم وجود حلول للمسألة.

الثالث: لا يمكن نَمْذَجةُ جميع المسائل ذات الأهمية في هذا الحقل في صيغة البحث الشجري.

- كذلك كانت «**حوسبةُ المنطق الرياضي**» [٣] سلاحًا ضاربًا أساسيًّا آخَرَ في ترسانة الذكاءِ الاصطناعي آنَذاكَ، ورغم النجاحات المبكرة التي حققها بالتوليد الآلي لإثباتات بعض «الـمُبَرْهَنات الرياضية theorems» – مما أغرى أسماءً كبيرةً في هذا الحقل بالزعم أن ذلك يمثل شوطًا كبيرًا في رحلة محاكاة الذكاء البشري بل

١– إذا افترضنا أنه في مسألةٍ ما – على سبيل المثال – كانت م = ٣ فإنه عند عمق ع = ٥ تكون ج = ١٢١ حالةً، وعند عـمق ع = ١٠ تكون ج = ٢٩٥٢٤ حالةً، أما عند عمق ع = ١٥ تتجاوز قيمة ج السبعة ملايين حالةً، وبالوصول إلى عمــق ع = ٢٠ تتجاوز قيمة ج ألف وسبعمئة وأربعين مليون حالة! وهكذا فإن أي حاسوب مهما عظمت قُدُراتُه سوف يعجز سريعًا مع تزايد عمق البحث، ولن يسعفنا حتى «قانون موورMoore’s law» (الذي ألمحنا إليه قرب نهاية القسم الفرعي ٤ ,١ من هذا الفصل) الذي يصف نمطَ زيادةِ قدرات الحواسيب مع مرور السنين.

٢– تُفِيدُ أفضل خُوارِزمات البحث العديدة (التي ألمحنا إلى بعضها آنفًا بعد الشكل رقم ٢) لترويض الانفجار الأُسِّيِّ الملازم للبحث الشجري في تقليل ذلك العدد الإجمالي بما يكافئ تقليل معامل توليد الحالات م، لكن نمو هذا الإجمالي يظل ذا نمطٍ أُسِّيٍّ مما يؤجل قليلاً استنفاد القدرات الحاسوبية مع زيادة عمق البحث لكنه لا يغير نمط النمو إلى نمطٍ آخر أهدأ.

٣– وتحديدًا «المنطق الرياضي من الدرجة الأولى Predicate Calculus» – وفي باطنه «المنطق الرياضي من الدرجة الصفرية» الذي يُعْرَفُ أيضًا بـ «المنطق الكلاميّ أو المنطق الأَرِسْطيّ Propositional Logic» كحالةٍ خاصةٍ بسيطةٍ منه – حيث يؤول التحليلُ المنطقيُّ الرياضيُّ المحوسَبُ إلى عمليةٍ تُسَمَّى «التفنيد refutation» التي تتكون بدورها من شِقَّيْنِ يسيران آنيًّا؛ ففي الشق الأول تتولد شجرة بدائل «التعبيرات المنطقية الرياضية predicates» ومساراتها، وفي الشق الثاني يتواصل البحث في هذه الشجرة أولاً بأول حتى الوصول إلى «العبارة المنطقية الخالية empty clause» حيث تنتهي عملية التفنيد حينئذٍ بأن العبارة المنطقية الرياضية المطلوب إثباتُها يصح ثبوتُها منطقيًّا من المسلَّمات الابتدائية.

وتخطيه كما أسلفنا – فإن هذا المنطق الرياضي المحوسَب يعاني أيضًا من ثلاثة عيوبٍ (على الأقل) يطعن كلٌّ منها بشدةٍ في نَجَاعتِهِ:

الأول: معضلة «التضخم الأسي exponential growth» في توليد شجرة التعبيرات المنطقية الرياضية، وهي المعضلةُ نفسُها التي ناقشناها في نقد البحث الشجري آنفًا.

الثاني: إذا انتهت عملية «التحليل المنطقي» – المسمَّاة أيضًا بـ «التفنيد refutation» – بالعثور على «العبارة المنطقية الخالية empty clause» فقد ثَبُتَت العبارة المنطقية المطلوب إثباتها وحُلَّت المسألة[1]، ولكن إذا لم يُعْثَرْ على تلك العبارة المنطقية الفارغة فلا سبيل إلى تقرير ما إذا كانت العبارة المطلوب إثباتها صحيحة أم خاطئة استنادًا إلى المسلَّمات الابتدائية. وبالتالي فلا ضمانةَ أن هذا الأسلوب سينتهي دومًا بالبت في صحة المسألة المطلوبة أم عدمها.

الثالث: لا يمكن نَمْذَجةُ جميع المسائل ذات الأهمية المعتبَرة في هذا المضمار على أنها تعابيرُ منطقيةٌ رياضيةٌ يُرادُ إثباتُها عبر آليةِ المنطقِ الرياضيِّ المحوسَب.

وينبغي عند هذه النقطة التنويه إلى أن هناك آلياتٍ وخوارزماتٍ أخرى لا يتسع المقام لسردها جميعًا وظفها الذكاء الاصطناعي في محاولة حل المسائل فيما سُمِّيَ بسنواته الذهبية من منتصف خمسينيات إلى مطالع سبعينيات القرن العشرين الميلادي، وهي آلياتٌ تشترك مع «البحث الشجري» و «المنطق الرياضي المحوسَب» في العيوب الثلاثة الأساسية التي ذكرناها، ولكننا مَثَّلْنا بهاتين العائلتين من الآليات الحاسوبية لكونهما الأبرز والأعم والأكثر شيوعًا فيما استخدمه الذكاء الاصطناعي في تلك الفترة. كما ينبغي التأكيد أيضًا على أن كل هذه الآليات الحاسوبية لا تزال مهمةً ومفيدةً حتى وقتنا هذا؛ حيث يجري تلافي عيوبها بالتقيد بالشروط الحاسوبية المناسبة للمسائل التي تعالجها، وبدمجها مع آلياتٍ حاسوبية أخرى سوف نتعرض لبعضها لاحقًا في هذا الفصل.

١- انظر الهامش السابق.

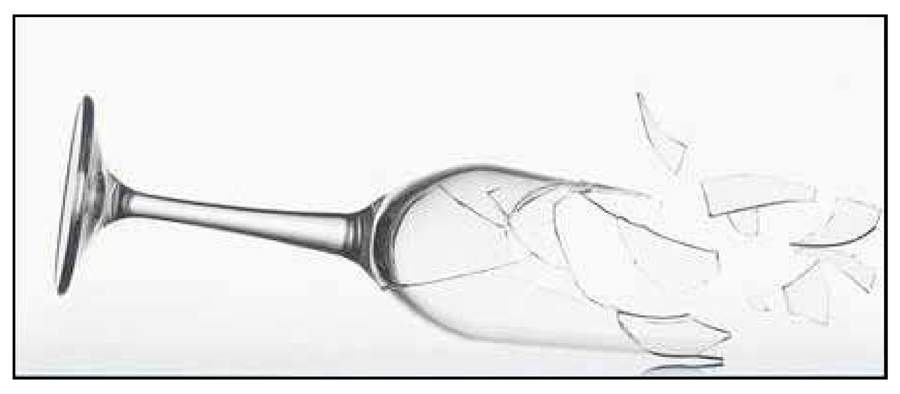
- بالوصول إلى منتصف سبعينيات القرن العشرين الميلادي اقتنعت جمهرة الباحثين بعد تجارب كثيرة أن إنجازَ تقدمٍ ملموسٍ نَحْوَ محاكاة المهام الإدراكية لدى الإنسان – مثل السمع: بمعنى تفسير الدماغ للذبذبات الصوتية الواصلة إليه من العصب السمعي، أو البصر: بمعنى تفسير الدماغ للإشارات الضوئية الواصلة إليه من العصب البصري، أو اللمس – ونَحْوَ اتخاذ الإنسان للقرار بناءً على هذا الإدراك يعتمد ضمن أمورٍ أساسيةٍ أخرى على معرفة البيئة الخارجية المحيطة به.

وربما كانت «مُفارَقة مُورافِيكْ»[1] من أكثر ما شاع في أدبيات الذكاء الاصطناعي تعبيرًا عن ذلك ومؤدَّاها أنه «على عكس الاعتقاد الـمُسْبَق الشائع أن عمليات التحليل والبرهنة المنطقية الرياضية تتطلب قدراتٍ حاسوبية أعلى من تلك التي تتطلبها العمليات الإدراكية الأساسية لدى الإنسان والحيوان، فإن الواقع هو أن محاكاة عمليةٍ إدراكيةٍ بسيطةٍ – كالتعرف على وجهٍ أو كالسير عبر غرفةٍ دون ارتطام بأثاثها – تحتاج أضعافَ أضعافِ الطاقات الحاسوبية التي تحتاجها عمليات التحليل والبرهنة المنطقية الرياضية الآلية»، ومَرَدُّ تلك الصعوبة معضلتان أساسيتان:

الأولى: الضخامة الهائلة للبيئة الخارجية؛ حتى لو كانت جزءًا محدودًا من العالم الواقعي، ويكفي لتصور تلك الضخامة إجراءُ محاولةٍ يائسةٍ لحَصْرِ ما يستوعبه دماغ طفلٍ صحيح في الثالثة من عمره عن العالم الخارجي، ناهيك عن عدم تجانس تلك المعرفة.

الثانية: تمثيل أغلب معرفة الإنسان عن العالم الخارجي في ذهنه تمثيلاً «كيفيًّا qualitative» لا تمثيلاً «كميًّا quantitative»، وكذلك استخدام آليات كيفية لا كمية في معالجة هذه المعرفة لإدراك متغيرات هذا العالم واتخاذ القرار، وتُسَمَّيان «المعرفة الفِطْرية commonsense knowledge» و«التحليل الفِطْري commonsense reasoning» على الترتيب، وما زالت حوسبتُها حتى الآن معضلةً عسيرةً (انظر الشكل ٤ أدناه والهامش المتصل به للمزيد من التمثيل والتوضيح).

١- الباحث الكَنَدي النِّمْساوي «هانْز بِيتَر مُورافِيكْ» هو أستاذٌ بمعهد الرُّوبوتيَّات في «جامعة كارْنِيغي مِيلُون» في «بِنْسِلْفانْيَا» في الولايات المتحدة الأمريكية، وهو ممن ارتبطت أسماؤهم بشدة بهذه القضية الأساسية في حقل الذكاء الاصطناعي.



الشكل ٤ : مثالٌ على المعرفة الفيزيائية الفِطْرية: كأسٌ زجاجيةٌ تسقط فترتطمُ بأرضيةٍ صلبةٍ فتتهشم[١]

- اتَّسَمَت الأساليبُ الأساسيةُ لمعالجة المسائل إبَّانَ «السنوات الذهبية» بكونها «قطعية/ تحديدية deterministic» بينما البياناتُ المدخَلةُ إلى تلك المسائل من البيئة المحيطة ذاتُ مُرَكّباتٍ «عشوائيَّةٍ stochastic» لا يمكن إغفالُها، بل إنه قد استقر في الرياضيات والعلوم التجريبية في العقود اللاحقة أن العشوائية عاملٌ متأصِّلٌ في العمليات والظواهر الطبيعية؛ فيما يُعْرَفُ بنظرية «الشَّوَاش chaos».

١ – تَصَوَّرْ حوارًا يقول فيه أحدُهم لصاحِبِه الذي يحمل كأساً زجاجيةً «احذَرْ أن تسقط من يدك الكأس.»، ثم تَصَوَّرْ ما الذي تستدعيه هذه العبارة البسيطة لحامل الكأس الزجاجي؟

إنها تستدعي: أن (أ) هناك مجال جاذبية نحو الأرض، وأن (ب) الكأس جِرْمٌ ذو كتلةٍ إذا فلت من يده سوف يتسارع نحو الأرض حتى يرتطم بأقرب سطح (الأرضية)، وكذلك أن (ج) الكأس مصنوعة من زجاج وهو مادة هشة إذا اصطدمت مسرعةً بسطح صلب فإنها على الأغلب ستتهشم، (د) وإذا تهشمت فإن أجزاءها ستكون ذات حوافَّ مدببةٍ، وأن (هـ) هذه الأجزاء ستتناثر بسرعة في اتجاهات عشوائية، وأن (و) أي جزء مدبَّب منها إذا اصطدم مسرعًا ببشرة الإنسان فمن السهل أن يخدش أو يقطع الجلد و اللحم وهما مادتان لينتان، (ز) وفي الغالب فإن ذلك سوف يسبب جرحًا يستتبع نزيفًا، وأن (ح) النزيف ربما ... إلخ.

يلاحَظُ بوضوح غزارة المعرفة التي استدعتها تلك العبارة البسيطة عن العالم الخارجي الواقعي، كما يُلاحَظُ بوضوحٍ أيضًا تسلسل التحليل المنطقي بسلاسةٍ للوصول إلى المخاطر المحتملة، وكذلكَ يُلاحظُ أن هذه المعرفة وتسلسلَ تحليلها المنطقي نَوْعِيَّان لم تُسْتَدْعَ فيهما قوانين الحركة لتكوين معادلات تفاضلية مثلاً يتوجب حلها للحصول على مسار الجسم الساقط وسرعة ومكان اصطدامه ثم حركة كل جزء من أجزائه المتهشمة ... إلخ مثلما يعالج علم «الفيزياء» عادةً مسائل الحركة معالجةً رياضيةً رصينةً، وكذلك يُلاحَظ السرعة اللحظية لاستدعاء هذه المعرفة الفِطْرية وإجراء تلك المعالَجة الفطرية ومن ثم اتخاذ حامل الكأس قرارًا بالحرص على عدم سقوط الكأس من يده وتنفيذ هذا القرار حَرَكيًّا لتفادي المخاطر المترتبة على سقوط الكأس من يده.

وحتى وقتنا هذا لم يتم تطوير أدوات رياضية ناضجة لحوسبة مثل هذه المعرفة الفِطرية عن العالم الواقعي (التي تُسَمَّى أيضًا «أونْطولوجْيَا ontology» العالم الواقعي) أو لإجراء مثل هذا التحليل الفِطري لها بهذه الكفاءة الرفيعة؛ مما يترك نقصًا فادحًا في ترسانة الذكاء الاصطناعي عمومًا ومعالجة اللغات الطبيعية على وجه الخصوص.

- في خمسينيات القرن العشرين الميلادي اكتشفت الدراسة التشريحية للمخ والأعصاب في البشر – وكذلك في الثدييات العُلْيا/ الراقية كالقِرَدة – أن المنظومة العصبية لتلك الكائنات الـمُعَقَّدة هي شَبَكَةٌ ذاتُ عدد هائل (يقدر بآلاف الملايين) من الخلايا العصبية يتصل كلٌّ منها بالعديد من الخلايا العصبية الأخرى في الشبكة. وفي العقود التالية تأكَّدَ هذا الاكتشافُ وأُضِيفَ إليه تحديدُ مواضع المراكز العصبية على هذه الشبكةِ كثيفةِ الترابطِ حيث يتعاظم النشاط الكهربي لكل مركزٍ مع أداء طائفةٍ معيَّنةٍ من الوظائف الذهنية أو الإدراكية للكائن الحي.

شَجَّعَتْ هذه الاكتشافاتُ باحثي الذكاء الاصطناعي على أن يحاولوا محاكاةَ هذه الشبكاتِ العصبيةِ محاكاةً حاسوبيةً مبسَّطةً [١] رجاءَ الحصول على محاكاةٍ للأداء الذهني أو الإدراكي للكائنات الحية المعقَّدة، فجَرَتْ محاولاتٌ أوليةٌ في حدود ما سمحت به القدرات الحاسوبية آنذاك واستطاعت إبداءَ تصرُّفٍ مشابهٍ لبعض العمليات «المنطقية الثنائية binary logic»، لكن برامج الأبحاث في هذا الاتجاه توقفت بنهاية الستينيات من القرن نفسه؛ ليس فقط بسبب محدودية القدرات الحاسوبية ولكن أيضًا بسبب موجةٍ عاليةٍ من الانتقادات للآفاق التي يمكن أن تصل إليها هذه المحاكَيَات [٢]. الأهم من كل ما سبق هو أن ذلك هو كُلُّ

١– سوف نعود للنماذج الحاسوبية للشبكات العصبية في القسم الثاني من هذا الفصل، كما سنتناولها مرةً أخرى بتفصيلٍ أكبرَ وبرسومٍ توضيحيةٍ في القسم الثالث منه كمثالٍ على الفارق بين واقع الظاهرة الطبيعية ونموذجها الرياضي في العلوم التطبيقية.

٢– كان «فرانْك روزِنبْلات» عالم النفس الأمريكي في «جامعة كورْنيل» في «نيويورك» بالولايات المتحدة الأمريكية هو رائد الدعوة لمحاكاة الشبكات العصبية حاسوبيًّا في عام ١٩٥٨م فانطلق برنامج أبحاث نشط لإجراء هذه المحاكاة وتطبيقاتها طوال عقد الستينيات، ولكن كل شيء توقف عام ١٩٦٩م حين نشر كُلٌّ من «مارْفِين مِينِسْكي» و «سايْمور بابِيرْت» كتابَهُما الشهير «خوارِزْمات الشبكات العصبية Perceptrons» اللذين قدما فيه حُجَجًا رياضيةً وفلسفيةً قويةً تبين انسداد الأفق أمام النماذج الحاسوبية للشبكات العصبية كسلاح فَعَّالٍ في جعبة الذكاء الاصطناعي.

استمر هذا التوقف ما يُقارِبُ عِقْدًا من الزمان حتى استؤنف العمل البحثي في هذا الاتجاه ثانيةً في عِقْد الثمانينيات حيث سيتبين رياضيًّا أن هذه النماذج الحاسوبية للشبكات العصبية أدواتٌ مفيدةٌ للغاية في «التمثيل التقريبي للدَّوَالِّ عديدة المتغيرات multivariate function approximation» بمجرَّد معرفة نماذج غزيرة من مدخَلاتها ومخرَجاتها ومن ثم ستصير أحد الوسائل الأساسية في «التعلُّم الحاسوبي machine learning» كما سيأتي لاحقًا في هذا الفصل.

وفضلاً عن أن «مينسكي» و «بابيرت» اسمان ذوا نفوذٍ كبير في هذا الحقل كائنين من الآباء المؤسسين له، فإن المفارَقةَ هي أنهما كانا أول من نَفَّذَ نموذجًا حاسوبيًّا لهذه الشبكات العصبية – وقد أوردنا ذلك في بدايات القسم الفرعي ١.٥ من هذا الفصل – كما أن هناك مفارَقةً شخصيةً طريفَةً أيضًا تتمثل في أن «مارفين مينسكي» كان زميلَ دراسةٍ لـ «فرانك روزينبلات» في المرحلة الثانوية ثم كان سببًا في تحطيم برنامجه البحثي الذي لم يُستأنف إلا بعد موته حيث توفي «روزينبلات» في عام ١٩٧١م.

ما نعرفه - حتى الآن - عن كيفية عمل الدماغ البشري والجهاز العصبي من حيث معالجته للمسائل الذهنية والإدراكية، وبالطبع فهو غير كافٍ أبدًا لمحاكاة أداء هذا الدماغ.

- الكثير من المسائل التي تُعْتَبَرُ مَحَكًّا حقيقيًّا لنَجَاعةِ الذكاء الاصطناعي - وعلوم الحاسب على وجه العموم - تبلغ من العُسْرِ والتعقيدِ مبلغًا يَصْعُبُ معه التعامل حاسوبيًّا مع أية نماذج رياضيةٍ لمكوناتها وآليات عملها، وتتجاوز الصعوبة ذلك أحيانًا فيتعذر من الأصل وضع نماذج رياضية قابلة للحوسبة لتلك المكوِّنات والآليات، بل تتصاعد الصعوبة أحيانًا أخرى فلا تُعْرَفُ عندئذٍ مكوِّناتٌ ولا آلياتٌ للمسألة [١].

ويمكن التمثيل لتلك المسائل بمسألة أساسية تتمثل في «التعرف على مفردات الكلام المنطوق Speech recognition»؛ فإذا تتبعنا مكونات وآليات عمل هذه الوظيفة الإدراكية لصادَفْنا ابتداءً الجهازَ الذي يستقبل ذبذبات الصوت وهو الأذن بأجزائها الكثيرة المركبة (الصُّوان، القناة السمعية، المِطْرَقة، السندان، طبلة الأذن، القنوات الهلالية، القوقعة، ... إلخ)، ثم العصب السمعي، ثم عمل المخ على الذبذبات التي يوصلها إليه العصب السمعي. والحقيقة أن محاولة وضع نموذج رياضي لبعض الأجزاء المنفردة من الأذن مَهَمَّةٌ عسيرة، أما وضعها مَعًا في نموذجٍ واحد فهو أمر في عِدَادِ الأحلام. فإذا انتقلنا إلى عمل المخ على تفسير الذبذبات التي تصل إليه من العصب السمعي فإن آليته غير معروفةٍ من الأساس! وإذا نظرنا في وظيفة التعرف البصري على الأشكال - كمثالٍ آخر - فإنها لا تقل صعوبة إن لم تزد عن السمع.

ويبقى المورد الوحيد الذي يمكن توفيره بيُسْرٍ لوصف مثل تلك المسائل هو نماذجُ غزيرةٌ من مُدْخَلاتِها بالتوازي مع مُخْرَجاتها، ومن ثَمَّ نبعت الحاجَة إلى ابتكار أساليب لمعالَجَةِ مثل تلك المسائل بالارتكاز على هذا المورد فقط، وبالفعل تمثَّلت الاستجابةُ فيما بعد في ميلادِ وتطوُّرِ عددٍ من الأساليب المتنوعة التي تلبي هذه الحاجة والتي تكوِّن إجمالاً ما يُعْرَفُ بمدرسة «التعلُّم الحاسوبي machine learning» [٢].

١- ويُطْلَقُ أحيانًا على مثل تلك المسائل اسم «المسائل غير المهَيْكَلة/ الهَيْكَلية unstructured problems».

٢- ستكون لنا وقفةٌ مع هذه المدرسة في مقارَبةِ المسائل الإدراكية والذهنية - وفي بؤرتها نمذجة ومعالَجة اللغات الطبيعية - في القسم الثاني وكذلك الثالث من هذا الفصل.

بَقِيَ أن نشير في ختام هذا القسم الفرعي إلى مشكلة أخرى أثارها الباحثون في نقدهم لحقبة «السنوات الذهبية» للذكاء الاصطناعي، ألا وهي محدودية قدرات الحواسيب آنذاك قياسًا بما يتطلبه حَسْبَ تقديرِهم حَلُّ بعض المسائل ذات الأهمية من موارد حاسوبية بنِسَبٍ تتراوح بين مئات المرات وآلاف المرات؛ فعلى سبيل المثال في عام ١٩٧٦م قدر الباحث «هانز بيتر مورافيك» - الذي سبق ذكره وترجمته في هذا القسم الفرعي - أن محاكاة أداء شبكية العين البشرية في اكتشاف وتتبع الأجسام المتحركة يحتاج ما بين ثمانين إلى ثمانية آلاف ضِعْف القدرة الحاسوبية لأقوى «حاسوبٍ فائقٍ» (سوبر كمبيوتر) في ذلك الوقت. ومن الطريف أَنْ نعرف أَنَّه مع نهاية عِقْد التسعينيات من القرن العشرين الميلادي كانت تلك القدرة الحاسوبية متوفرةً بسهولة في العديد من الحواسيب القوية التي كانت متاحةً في الأسواق وقتذاك بأسعار لا تتجاوز خمسة آلاف دولار أمريكي. ويعني ذلك للوهلة الأولى أن مشكلة محدودية القدرات الحاسوبية كانت وقتيةً، وأن الزمن كان كفيلاً بحلها حيث إنه كما ذكرنا في القسم الفرعي ١,٤ من هذا الفصل فإن هناك نمط تزايدٍ مُرَكَّبٍ مطَّرِدٍ لتنامي القدرات الحاسوبية مع الزمن يشتهر باسم «قانون موور» ومؤداه أن تلك القدرات (من حيث عدد التعليمات المنفَّذة في الثانية الواحدة، وسعة الذاكرة) تتضاعف تقريبًا كل ١٨ شهرًا [١]؛ فمثلًا في عام ٢٠١٥م وبعد تسعة وثلاثين عامًا من عام ١٩٧٦م تضاعفت القدرة الحاسوبية بمقدار اثنين مرفوعةً لأس ستة وعشرين [٢] أي حوالي سبعة وستين مليون مرةً !

ورغم أن هذا التزايد فائق التسارع مع الزمن للقدرات الحاسوبية يبدو بَرَّاقًا داعيًا لتفاؤلٍ كبيرٍ، فإن الحقيقةَ مراوِغةٌ بعض الشيء حيث إن القدرات الحاسوبية ليست فقط هي التي تتزايد باطِّرادٍ ولكن تتزايد معها أيضًا التحديات التي تمثِّلُها المسائلُ المطلوب حَلُّها نوعًا وحجمًا.

١- يجب الانتباه إلى أن نمط التزايد الأُسِّيِّ للقدرات الحاسوبية لا يمكن أن يَطَّرِدَ مع الزمن إلى الأبد، وذلك لأن هذا التزايد الذي استمر عبر العقود الستة الماضية يرتبط بالتصغيرِ المُطَّرِدِ للوحدات الإلكترونية (الترانزيستورات) وهو ما قارب على الاصطدام بحدودٍ فيزيائيةٍ لا يمكن تخطيها.

٢- لاحظ أن ٢٦ × (٣ ÷ ٢) = ٣٩.

٨, ١ . الانبعاث وإعادة النهوض

كانت «النُّظُم الخبيرة expert systems» أَحَدَ المساراتِ الأقلِ لَمعاناً التي كان يعمل عليها بهدوءٍ بعض باحثي الذكاء الاصطناعي منذ منتصف ستينيات القرن العشرين الميلادي؛ حيث يتكون «النظام الخبير» من مُكَوِّنَيْنِ أساسيين؛ أولهما هو قاعدة واسعة قدرَ الإمكان من الحقائق حول موضوعٍ محدَّدٍ ضَيِّقِ النِّطاق (مثل: ري الخَضْرَاوات بالتنقيط في الصَّوْبات الزراعية) المستقاة من خُبَرَاءَ في هذا المجال والمصوغة في قوالبَ منمَّطةٍ تركيبيًّا واصطلاحيًّا، وثانيهما هو آليةٌ للاستدلال والاستنتاج عبر التحليل المنطقي الرياضي[١] تعمل على المكوِّن الأول: «قاعدةِ حقائق الخبرة». وعند تَلَقِّيهِ سؤالًا في نطاق الموضوع الذي يغطيه، فإن النظام الخبير عبر التحليل المنطقي الرياضي لـ«قاعدة حقائق الخبرة» المتوفرة لديه قد يُوجِّه بدوره إلى المستخدِم الذي وجه إليه السؤال الأساسي بعضَ الأسئلة التكميلية ذات الأجوبة البسيطة («نعم أم لا»، أو «اختيارٍ من متعددٍ») ويعتمد توليد كل سؤال تكميلي في هذا الحوار على إجابات سابِقِيهِ، ويستمر ذلك إلى أن تتوصل آلية الاستدلال والاستنتاج إلى إجابةٍ للسؤال الأصلي[٢].

ومن النجاحات المبكِّرة التي أظهرت فَعَاليةَ وعَمَلانِيَّةَ النظم الخبيرة نظامٌ سُمِّيَ «دِنْدِرال Dendral» الذي بدأ تطويره عام ١٩٦٥م بواسطة «إدوارد فِيغِينْبَاوْم»[٣] وطلابُه حيث حاكَى النظام دور الخبير الكيميائي بتحديد المركبات الكيميائية عبر دراسة «تحليلها الطيفي spectrograph»، وكذلك نظامٌ آخر سُمِّيَ «مايْسِين MYCIN» اكتمل تطويره عام ١٩٧٢م واستطاع عبر دراسة نتائج تحليل الدم تشخيص الأمراض المعدية عن طريق نقله من شخص إلى آخر.

أما النجاح الفارِق للأنظمة الخبيرة الذي أعاد طرحَ «الذكاء الاصطناعي» في أوساط الجهات الحكومية والصناعية وفي الإعلامِ واسعِ الانتشار كحقلٍ علميٍّ وبحثيٍّ جادٍّ بعد «سنواته العِجاف» – انظر القسم الفرعي السابق – فقد أتى عام ١٩٨٠م عَقِبَ

١– والذي تناولناه في القسمين الفرعيين السابقين ٦, ١ و ٧, ١.

٢– أو إلى أن يستنفد التحليلُ المنطقي فيها الحدَّ الأقصى للموارد الحاسوبية المتاحةِ له وعندئذٍ يُعلَن عَدَمُ العثور على حَلٍّ.

٣– وكان أستاذًا لعلوم الحاسب وباحثًا في الذكاء الاصطناعي بجامعة ستانفورد بالولايات المتحدة الأمريكية، ويُطْلَق عليه لقب «الأب المؤسس» للأنظمة الخبيرة.

اكتمال تطوير نظامٍ أُطْلِقَ عليه اسمُ «إكس كون XCON» في جامعة «كارنيغي ميلون» الأمريكية لصالحِ شركة أنظمة الحواسيب الرقمية الكبرى وقتها «مؤسسة المعدات الرَّقْمية Digital Equipment Corporation (DEC)».

فقد سَجَّلَ هذا النِّظامُ الخَبيرُ أوَّلَ نجاحٍ صناعيٍّ واقتصاديٍّ ذي بالٍ للذكاء الاصطناعي؛ حيث أوردت الشركة في تقاريرها الرسمية عام ١٩٨٦م أن هذا النظام يوفر للشركة سنويًّا أربعين مليون دولارٍ أمريكي - وتكافئ قيمتها الشرائية عام ٢٠١٩م على أقل تقدير خمسةً وتسعين مليون دولار أمريكي. وينبع هذا الوفر من استغناءِ كل موظَّف مبيعاتٍ لدى المؤسسة بهذا النظام عن الحاجة إلى خَلِيَّةٍ من الفنيين في هندسة الحواسيب تحدد مكونات «العَتَاد hardware» و «البرمجيات software» الصحيحة والمتوائمة معًا التي تلبي بالضبط طلب كل عميل دون نقصٍ أو هَدْرٍ [١].

ومما يلفت النظر أن الأنظمة الخبيرة مثلت اتجاهًا عكس التيار السائد وَقْتَها في الذكاء الاصطناعي الذي كان ينحو إلى الاعتماد كأقصى ما يمكن في معالجة المسائل على التحليل الرياضي والخوارزمات، فيما ينحو أيضًا قدر المستطاع إلى اختزال البيانات والمعلومات المدخَلة إلى عمليات المعالجة تلك. وربما كان ذلك الاتجاه من الخطوات الأولى على الطريق الذي أوصل الذكاء الاصطناعي في السنوات الأخيرة إلى نضوجِ مقارَبة العديد من المسائل الهامة عبر «التعلُّم الحاسوبي» وعبر معالجة «البيانات العملاقة big data».

ويجري بناءُ «قاعدة حقائق الخبرة experience facts base» لأي «نظامٍ خبيرٍ» باستقائها من العاملين ذوي الخبرة الراسخة في نطاق عمل هذا النظام عبر مقابلاتٍ مطوَّلةٍ معهم يجريها مطورو النظام الذين يطرحون عليهم الأسئلة بأنساقٍ مصمَّمةٍ

١- لم تكن الأنظمة ذات القدرات الحاسوبية العالية وقتذاك التي تبيعها المؤسسة تأتي كباقةٍ واحدةٍ معيارية يختارها العميل من بين عدة نماذجَ محدودةٍ سابقة التجهيز والتجميع تطرحها الشركةُ في الأسواق - كما هي العادة الآن - بل كان العملاء يصفون احتياجاتهم الحاسوبية لموظَّفي المبيعات وبِناءً على هذا الوصف تتحمل الشركة عبء انتقاء العشرات - وربما المئات - من مكوِّنات العتاد والبرمجيات اللازمة لتلبية هذه الاحتياجات ومن ثم تقوم بتجميعها على هيئة نظام متكاملٍ مفصَّلٍ خِصِّيصًا لكل عميل. ومع ملاحظة أن تكلفة كل عنصرٍ من عناصر العتاد أو البرمجيات كانت باهظةً جدًّا بالمقارنة بأسعار وقتنا الراهن - بسبب إنتاجها بكميات محدودة على نطاق ضيق - فإن إضافة عناصر لا يحتاجها العميل كانت تمثل خسارةً ماليةً مؤلمة.

لتغطية مختلف السيناريوهات المحتمَلة للحوار وكذلك لكشف أي تناقض في إجاباتها، كما يضعون إجاباتهم في قوالبَ مصممةٍ للسماح بتحليلها تحليلاً منطقيًّا رياضيًّا. وكي لا تقف معضلة «الضخامة الهائلة للمعرفة الفطرية» - التي تناولناها في القسم الفرعي السابق - سدًّا منيعًا يحول دون بناء «قاعدة حقائق الخبرة» فإن كل «نظام خبيرٍ» يقيد نفسه بنطاق معرفيٍّ ضيقٍ للغاية على غرار المثالين اللذين ذكرناهما آنفًا. [١]

وعلى صعيدٍ آخرَ أثناء سنوات الانبعاث ذاتِها - وهي السنوات السبع الأُوَل من ثمانينيات القرن العشرين الميلادي - وقعت تطوراتٌ هامةٌ أعادت إحياء المحاكاة الحاسوبية للشبكات العصبية بعد اضمحلال العمل عليها بنهاية ستينياتِ القرن نفسِهِ - كما جاء في القسم الفرعي السابق - فقد ابتكر الفيزيائيُّ الأمريكيُّ «جون هُوبْفِيلْد» عام ١٩٨٢م [٢] نموذجًا جديدًا من الشبكات العصبية الحاسوبية وبرهن رياضيًّا أنها تستطيع لعب دور «الذاكرة الارتباطية associative memory» التي تمثل مَلْمَحًا مميِّزاً للذاكرة البشرية مقابل «ذاكرة الاستدعاء المباشر random access memory» المميِّزة للحواسيب الرقمية المعيارية [٣]. وفي الوقت نفسه تقريبًا نجح «جِيفْري هانْتُونْ» [٤] مع «دِيفِيدْ رُومِيلْهارْت» [٥] في البناء على الأسس الرياضية التي أرساها «سِيبُّو لِينَايِنْمَا» [٦]

١- وبِلُغَةِ الرياضيات التي تصف معضلات بناء وتمثيل ومعالجة «المعرفة الفطرية» فإن «قاعدة حقائق الخبرة» التي يرتكز عليها أيُّ «نظام خبيرٍ» هي شريحةٌ أونطولوجيةٌ ضئيلة الحجم للغاية - لكنها مفصَّلةُ التمثيل - من الأونطولوجيا الكلية للعالم الواقعي.

٢- وكان أستاذًا في «معهد التقنية بجامعة كاليفورنيا (كالْتِيكْ) Caltech» وهو كذلك أحد مؤسسي برنامج الدكتوراة في «الحوسبة والشبكات العصبية» في المعهد ذاته.

٣- من أجل تخزينها في «ذاكرة الاستدعاء المباشر» فلا بد أن تُعْطَى كلُّ معلومة «عنوانًا مفهرسًا index» ومن أجل استرجاع تلك المعلومة فلا بد كذلك من معرفة هذا «العنوان المفهرس» وهذا هو النمط الذي تعمل وفقًا له ذاكرة الحاسوب الرقمي المعياري، أما الذهن البشري فلربما تتداعى إلى ذهنه فكرةٌ/ معلومةٌ واحدة أو أكثر (مثل: «المزروعات»، «الخُضْرة»، «النماء»، «الفيضان»، ...) عندما تُذْكَرُ له فكرةٌ/ معلومةٌ أخرى مرتبطةٌ بها (مثل: «سقوط الأمطار») ويسمى هذا النوع من الذاكرة التي تُستدْعَى فيها المعلومة/ المعلومات بمعرفة معلومةٍ أخرى مرتبطةٍ بها «الذاكرةُ الارتباطية».

٤- وهو أستاذٌ كَنديٌّ إنغليزيٌّ بارزٌ في «علم النفس الإدراكي» و «علوم الحاسب» في «جامعة تورونتو» في كندا ثم باحثٌ رئيسيٌّ في «شركة غووغِل» العملاقة الشهيرة للبرمجيات.

٥- وهو أستاذ أمريكي في «علم النفس الإدراكي» في «جامعة ستانْفُورْد» ثم في «جامعة كالِيفورْنِيا - سان دِيِيغُو» الأمريكيتين.

٦- وهو عالم الرياضيات الفنلندي والباحث في علوم الحاسب في «جامعة هِيلْسِنْكِي» في فنلندا، ثم الأستاذ الزائر في «جامعة مَارِيلانْد» بالولايات المتحدة الأمريكية.

في بداية عِقْد السبعينيـــات ومن ثَمَّ إحكـــامِ صياغةِ خُوارِزْم «الانتشار الارتدادي back propagation» لتدريب «الشبكـــــات العصبية متعـــــددة الطبقات multi-layer neural nets» - وهو النموذج الأصلي للشبكات العصبية المحوسَبة - على مدخلَات ومخرَجات ظاهرةٍ ما من أجل محاكاة سلوكها[1]؛ وهو التطور الأهم في مسيرة هذه الشبكات التي سوف تترسخ مع منتصف عِقْدِ التسعينيات على مُنْحَنًى تصاعديٍّ وصولاً إلى وقتنا الراهن كأحد الأسلحة الماضية في ترسانة «التعلم الحاسوبي machine learning» وهو ما سيلعب معنا دورًا هامًّا فيما تبقى من هذا الفصل.

بعثت هذه النجاحاتُ الآمالَ بقوةٍ بعد خُفوتها في الذكاء الاصطناعي من بابٍ عُرِفَ باسم «هندسة المعرفة Knowledge Engineering» فدفعت حالة النجاح الاقتصادي الذي حققته «النظم الخبيرة» - مثل حالة «إكس كون XCON» التي أسلفنا تفصيلَها - الكثيرَ من الشركات الكبيرة حول العالم في مجال الصناعة والخدمات إلى الإقبال على إدماج «النظم الخبيرة» في عملياتها بل إلى الاستثمار السخي في إنشاء أقسامٍ داخلها لأبحاث الذكاء الاصطناعي لتطوير «نُظُمٍ خبيرةٍ» خاصة بها، ومن أجل تلبية طلبات هذه الشركات وأقسام الذكاء الاصطناعي فيها نشأ كذلك قطاعٌ صناعيٌّ يشمل شركاتٍ للعتاد الحاسوبي الرقمي (مثل Symbolics و Lisp Machines) وشركاتٍ للبرمجيات (مثل IntelliCorp و AIon).

أما على صعيد الكيانات الأكبر فلم يكن انبعاث الدعم لأنشطة أبحاث وتطوير الذكاء الاصطناعي بأقلَّ من ذلك؛ ففي عام ١٩٨١م خصصت وزارة الصناعة والتجارة الدولية اليابانية ثمانمئةً وخمسين مليون دولار أمريكيّ لما أطلقت عليه وقتها «مشروع الحواسيب من الجيل الخامس» الذي تمثلت أهدافه في تطوير برمجيات وصناعة حواسيب تستطيع «إجراء حوارات باللغة الطبيعية» و «الترجمة بين اللغات الطبيعة» و «تفسير الصُّوَر» و «القيام بالتحليل المنطقي» بمستوى الأداء البشري، وقد اختيرت لغة «برولوغ Prolog» [2] لُغَةً أساسيةً لكتابة البرمجيات في هذا المشروع. وكذلك

١- أي حساب قيمة «أيٍّ من مخرَجات الظاهرة» المناظرة لـ «أيٍّ من المُدْخَلات إليها» من خارج عينة بيانات التدريب.

٢- وهي كما أسلفنا في القسم الفرعي ١.٦ من هذا الفصل لغةٌ لبرمجة الحواسيب الرقمية تُيَسِّر لمطوري البرمجيات استدعاء وتوظيف أدوات المنطق الرياضي التعريفية والتحليلية في برمجياتهم بسلاسةٍ.

أنفقت الحكومة البريطانية ثلاثمئةً وخمسين مليون جنيه إسترليني بين عامي ١٩٨٣م و ١٩٨٧م على مشروعٍ مشابهٍ تحت اسم «آلْفايْ Alvey»، كما تشكَّل تكتُّلٌ من شركات أمريكية كبيرة أُطْلِقَ عليه «تحالف الإلكترونيات الدقيقة والتقنيات الحاسوبية MCC» لتمويل المشروعات الضخمة في حقل الذكاء الاصطناعي وتقنيات المعلومات. أما آخر مثال نورده في هذا الصدد فهو «المبادرةُ الاستراتيجية للحوسبة» التي أطلقتها «وكالة مشروعات الأبحاث الدفاعية المتقدمة (دارْبَا) DARPA» في الولايات المتحدة الأمريكية بين عامي ١٩٨٤م و ١٩٨٨م [١] وعودتُها – بعد انقطاعٍ – للاستثمار في أبحاث الذكاء الاصطناعي مع مضاعفة استثمارها فيه ثلاث مرات.

لعل القارئ بعد هذا الانبعاث يترقب إعلان النهاية السعيدة لهذا التاريخ الطويل باستقرار مسيرة الذكاء الاصطناعي وازدهارها المُطَّرِد حتى وقتنا الراهن، ولكن الواقع هو أن موجةً أخرى من الجفاف والسنوات العجاف كانت على وشك أن تزحف على هذا الحقل والباحثين فيه!

ففي عامَيْ ١٩٨٦م و ١٩٨٧م تعاظَمَ إعجابُ المؤسسات التجارية والصناعية والهيئات الحكومية في العالم الغربي بالنُّظُمِ الخبيرةِ والإقبالُ على استخدامها والاستثمارُ فيها بتسارعٍ مُطَّرِدٍ، إلا أنه على الرغم من نجاح تلك النُّظُم في «التعلُّم الأوَّليّ» من الخبراء في تخصُّصٍ معيَّنٍ عبر استخلاص وحوسبة معرفتهم لاستخدامها عبر التحليل بأدوات المنطق الرياضي في استنتاج إجاباتٍ لمسائلَ عمليةٍ في النطاق «الضيق» لهذه المعرفة، فقد تَبَدَّت مع طول الاستخدام نقاطُ ضعف هذه النظم؛ وأهمها الصعوبة والتكلفة العاليتان لتعديل «قاعدة حقائق الخبرة» مما يعني ضعف القدرة على تحقيق «التعلم المتواصل» سواء لتحسين الأداء أو للتكيف مع التغيرات في الظاهرة التي تغطيها هذه القاعدة المعرفية، وكذلك هَشَاشةُ أداء هذه النظم حيث تكاد تكون إجاباتُها خَبْطَ عَشْواء مع تصديها لأسئلةٍ تخرج ولو خروجًا طفيفًا عن النطاق الضيق لقاعدتها المعرفية مما قد يؤدي إلى كوارث إذا اعتمدت العمليات الصناعية والخدمية الحيوية على تلك «النظم الخبيرة».

١– وذلك لخدمة المبادرة العسكرية الكبرى التي كان يروج لها بقوةٍ الرئيس الأمريكي «رُونالْد رِيجان» وسَمَّاها «حرب النجوم» في إطار جهود إدارته لحسم الصراع العالمي الطويل مع «الاتحاد السُّوفْيِيتِّيّ» عبر التفوق التِّقَنِيِّ السَّاحِقِ عليه، وبالفعل سقط «الاتحاد السوفييتي» وتفكك سياسيًّا مع ختام عام ١٩٩١م بعد استنزافٍ اقتصاديٍّ وعسكريٍّ كانت هذه المبادرة آخر مراحله.

وبختام عام ١٩٨٧م وقع انقلابٌ في عالم الحواسيب الرقمية حين بزغ عصر «الحواسيب الشخصية Personal Computers» من شركَتَيْ «آبِل Apple» و «آي بي إم IBM» التي تجاوت وتتذاك قدراتها الحاسوبية قدرات الحواسيب المتخصصة في الذكاء الاصطناعي (مثل Lisp machines من صناعة شركة Symbolics) في حين لم يبلغ سعرُ هذه الحواسيب الشخصية إلا نسبةً بسيطةً من سعر تلك الحواسيب المتخصصة، فتبخرت بين عشيةٍ وضُحاها خمسمئةُ مليون دولار أمريكي مع انهيار صناعة تلك الحواسيب المتخصصة وتطوير مستلزماتها البرامجية (من أنظمة تشغيل وخلافه).

وفوق ذلك في عام ١٩٨٩م ارتأت إدارةٌ جديدةٌ لـ «وكالة مشروعات الأبحاث الدفاعية المتقدمة (دارْبَا) DARPA» الامريكية أن «الذكاء الاصطناعي» ليس هو الموجة التقنية الجديدة الواعدة التي يجب أن تدعمها الوكالة وأن عليها أن تعيد توجيه ميزانيتها إلى مشروعات أخرى ذات مردوداتٍ أكثرَ واقعية، فكان بالفعل أن قطعت الوكالة دعمها بقسوةٍ عن مشروعات الذكاء الاصطناعي مرةً أخرى وأوقفت «مبادرتها الاستراتيجية للحوسبة».

ولم يكن الوضع بأفضل حالاً في اليابان؛ فمع الوصول إلى عام ١٩٩١م لم تتحق قائمة الأهداف مفرطة الطموح لـ «مشروع حواسيب الجيل الخامس» بعد عشر سنوات من انطلاقه.

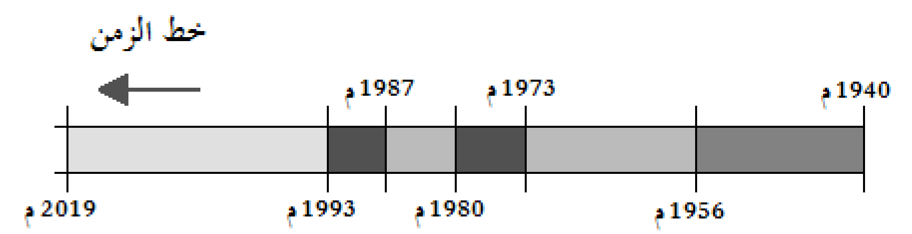
ولما كان الإعلامُ وأوساطُ الأعمالِ وبعض الشخصيات الأكاديمية قد بالغت مَرَّةً أخرى في سنوات الثمانينيات من القرن العشرين في النفخ في حدود قُدُراتِ وآفاق «الذكاءِ الاصطناعيِّ» بعامَّةٍ و «النظمِ الخبيرةِ» على وجه الخصوص، فإن تلك المبالغات قد أدت مع مُضِيِّ الوقت إلى خلق فُقَاعةٍ اقتصاديةٍ وتقنيةٍ كبيرة انفجرت بالفعل بنهاية ذلك العِقْدِ نَفْسِهِ جَرَّاءَ ما أسلفنا التمثيل له لتَوِّنا من نقاط قصور وتطورات تقنية وتمويلية؛ فبالوصول إلى عام ١٩٩٣م كان عدد ما اختفى من شركات تخصصت في الذكاء الاصطناعي - أو اعتمد نشاطُها عليه - ثلاثمئةَ شركةٍ أنشئت كلُّها في أعوام الانبعاث السبع بين ١٩٨١م و١٩٨٧م.

ولأنها المَرَّةُ الثانيةُ التي يتكرر فيها انزلاقُ مسار «الذكاء الاصطناعي» إلى أزمةٍ عامَّةٍ فلا بُدَّ لنا من تقريرِ وإبرازِ الدرس المستفاد التالي من تلك الخبرة التاريخية القيمة:

قاعدة رقم ١
قد تؤدي المبالغةُ في قُدُراتِ أي فرعٍ مستحدَثٍ من العلوم و/ أو التقنية، والتسرعُ في توسيع آفاقها التطبيقية بلا قيود، ووضعُ مستهدَفاتٍ مفرِطةٍ في التفاؤل؛ قد يؤدي كل ذلك إلى مكاسب إعلامية وتمويلية قصيرة المدى، لكنه من المؤكد أنها جميعًا تؤدي على المديين المتوسط والطويل إلى أثرٍ عكسيٍّ حتميٍّ يتمثل في فقدان الثقة في هذا المسار من العلوم والتقنية وتعطيله لفترة طويلة.

وبطبيعة الحال فلا بُدَّ كذلك من الاستفادة من هذا الدرس القيم وتطبيقه على حوسبة اللغات الطبيعية ومن بينها حوسبة اللغة العربية التي هي بؤرة هذا الكتاب.[١]

وقبل متابعة هذه المسيرة في مرحلتِها الأخيرة الممتدة إلى وقتنا الراهن، فمن المفيد أن نلخص جميع مراحلها على خارطة زمنية يعرضها فيما يلي (الشكل ٥)؛ حيث يفصل كل خط رأسي بين مرحلتين، وتتناسب درجة سطوع التظليل في كل مرحلة مع درجة النشاط والنضج فيها.



الشكل ٥: خارطةٌ زمنيةٌ معاصرةٌ مبسَّطةٌ لصعود وهبوط النشاط البحثي في حقل الذكاء الاصطناعي

١- وعلى الرغم من شدة وضوحها فإن مخالفة هذه القاعدة الذهبية لا يزال يتكرر مرةً بعد أخرى عند تفعيل مبادراتٍ طموحةٍ لحوسبة اللغة العربية في العالم العربي مسبِّبةً ليس فقط فشلَها بل أيضًا تشويهَ سمعة حقل حوسبة اللغة من حيث الجِدية والفعالية مما يثبط أيةَ نوايا مستقبليةً لتكرار المحاولة، وندعو القارئ كتمرينٍ عمليٍّ إلى البحث عن أمثلةٍ معاصرةٍ لمثل تلك الحالات.

فبالترحال على هذه الخارطة الزمنية نصادفُ:

أولاً: المرحلة التمهيدية التي تمتد من أربعينيات القرن العشرين الميلادي مع تبلور «علوم الحاسب» وبناء «الحواسيب الرقمية» الأولى حتى مؤتمر «دارتموث» الذي يعد بمثابة «الميلاد الأكاديمي» لحقل «الذكاء الاصطناعي» عام ١٩٥٦م.

ثانيًا: «السنوات الذهبية« التي امتدت إلى غاية عام ١٩٧٣م.

ثالثًا: الموجة الأولى من «السنوات العجاف» - أو كما يطلق عليها بعض المؤرخين «الشتاء الأول للذكاء الاصطناعي» - التي امتدت لغاية عام ١٩٨٠م.

رابعًا: «سنوات الانبعاث وإعادة النهوض» التي استمرت لغاية عام ١٩٨٧م.

خامسًا: الموجة الثانية من «السنوات العجاف» - أو كما يطلق عليها بعض المؤرخين «الشتاء الثاني للذكاء الاصطناعي» - التي امتدت لغاية عام ١٩٩٣م.

سادسًا: «الحقبة المعاصرة» الممتدة من ١٩٩٣م حتى وقتنا الراهن، والتي يطَّرِدُ فيها التطور البحثي والإنجاز التطبيقي بعد استقرار عدد من المفاهيم والمقارَبات الرياضية الحاسوبية في قناعة جمهور الباحثين في هذا الحقل - كما ستُبَيِّنُ الوُرَيقاتُ التالية - مما يجعلها جديرةً أن تُسمَّى أيضًا «حقبة البيانات العملاقة والتعلُّم الحاسوبي».

وقبل الولوج إلى «الحقبة المعاصرة» لا بُدَّ لنا من تقريرِ وإبرازِ قاعدة ذهبيةٍ أخرى[1] مستفادةٍ من هذه الخارِطة الزمنية:

١- وفي الحقيقة فإن عدم التقيُّد بهذه القاعدة الذهبية يشكل للأسف مشكلةً كبيرةً في الواقع الراهن لحوسبة اللغة العربية؛ فكثيرٌ من السادة اللسانيين العرب مشبعون بقناعاتٍ «منسوخةٍ outdated» ترجع إلى «مراحل باكرة» من تطور الذكاء الاصطناعي ومعالجَته لمَيْكَنةِ اللغات الطبيعية سابقةٍ على «حقبته المعاصرة». والسبب في ذلك هو الوقت اللازم كي تنتقل المفاهيم والمكتشَفات والمنجَزات التجريبية» الجارية في حقل الحوسبة إلى أدبيات اللسانيات الغربية إضافةً إلى الوقت اللازم بعد ذلك كي تُتَرْجَمَ إلى أدبيات اللسانيات العربية فضلاً عن الوقت اللازم ليهضمَها السادة اللسانيون العرب، ويُشَارُ إلى ذلك بوجود «فارق في الطَّوْر phase shift» بين اللساني العربي النابه الذي يعمل (في العقد الثاني من الألفية الميلادية الثانية على سبيل المثال) بقناعاتٍ قديمةٍ منسوخةٍ من حقل الذكاء الاصطناعي والحوسبة (نشأت في ثمانينيات القرن الميلادي الفائت على سبيل المثال).

فعلى سبيل المثال لا يزال عدد من اللسانيين العرب النابهين الآن يجاهدون منذ سنين طويلةٍ دون طائل في بناء نظم لمعالجة اللغة العربية آليًّا (كالترجمة الآلية من/ إلى اللغة العربية على سبيل المثال) بمقارَبات «النحو التحويلي» و «النحو التوليدي» ... وما إلى ذلك من أساليبَ تحليليةٍ أقلعت عنها منذ سنواتٍ طويلة أنجحُ نظم حوسبة اللغات الطبيعية المعاصرة. والأوفقُ من أجل تجاوز مشكلة «فارق الطور» تلك هو أن يعمل اللسانيون العرب النابهون في فرقٍ تضم إليهم باحثين نابهين في علوم الحاسب والذكاء الاصطناعي ولا سيما «التعلم الحاسوبي machine learning» كما سنبين لاحقًا في هذا الفصل.

قاعدة رقم ٢
على الباحث أو المطوِّر أو المخطِّط أو مدير المشروع في حقل حوسبة اللغات الطبيعية – ومنه «حوسبة اللغة العربية» التي هي بؤرة اهتمام هذا الكتاب – الحذرُ والتبصُّرُ الشديدين عند الرجوع إلى تراث وأدبيات الذكاء الاصطناعي قبل «الحقبة المعاصِرة» حيث كان التقلب في الرؤى والقناعات شديدًا في مراحله الباكرة، فإذا ما ابتغى أفضلَ الممارسات التي استقرت في هذا الحقل فعليه بالرجوع إلى المصادر الأحدث في الحقبة المعاصرة.

وعودةً إلى ما بعد انحسار الموجةِ الثانية من سنواته العِجاف التي امتدت قُرابةَ الست سنوات، فإنه يُنْظَرُ إلى مسيرة الذكاء الاصطناعي منذ عام ١٩٩٣م حتى وقتِنا الراهن كمرحلةٍ متصلةٍ واحدةٍ؛ وفيما يلي السمات العامة التي تميز هذه المرحلة:

- في المراحل السابقة كان هناك مَيْلٌ واضحٌ لدى باحثي الذكاء الاصطناعي إلى الاقتصار في الأدوات التي يستخدمونها على تلك التي طُوِّرَت تحت مظلة «الذكاء الاصطناعي»، وكذلك مَيْلٌ إلى تمييز عملهم وحقلهم عن سواه من حقول البحث والتطوير حتى تلك المتجاورة معه المرتبطة به ارتباطًا «عائليًّا» وثيقًا كالفروع الأخرى لعلوم الحاسب والرياضيات، و «علم الاتصالات Telecommunications»، و«معالجة الإشارات الرقمية Digital Signal Processing (DSP)»، و «بحوث العمليات Operations Research»، و «نظرية التحكُّم Control theory» ... إلخ. وفي المقابل تمتاز الحقبة المعاصرة بانفتاح بواباتها للاستقاء من روافد كل تلك الحقول [1] حتى إنه ليصعب الآنَ تعيُّن حدودٍ فاصلةٍ بين حقل الذكاء الاصطناعي وتلك الحقول الأخرى التي ترفده.
- غَلَبَ على الأساليب التي ارتكزت عليها معالَجاتُ الذكاء الاصطناعي للمسائل في المراحل السابقة أنها أساليبُ «قَطْعيةٌ/ تحديديةٌ deterministic» وقد أسلفنا

---

١ – فمثلاً: قدمت «معالجةُ الإشارات الرقمية» حلولاً عمليةً « للتعرُّفِ على الكلام المنطوق» و«التعرُّفِ على الأشكال المرئية».

أن ذلك نقطة ضعف كبيرةٌ [1] حيث إن البياناتِ المُدْخَلةَ إلى تلك المسائل من البيئة المحيطة ذاتُ مركّبات عشوائية واضحة فضلاً عن تأصل العشوائية والتعقيد في العمليات والظواهر الطبيعية. أما في هذه المرحلة فقد التحق بترسانة أساليب معالَجة الذكاء الاصطناعي للمسائل مَدَدٌ حافلٌ من أسلحة الرياضيـــات الماضية من «التحليل العَــدَدي Numerical Analysis» و «نظـــرية الاحتـــمالات Probability Theory» و «التحليـــــل الإحصـــــــائي Statistical Analysis» و «العمليــــــات اللا يقينية Stochastic Processes» و «نظـــرية الشّـــوَاش Chaos Theory» ... إلخ مما غير المشهد من المثالية إلى الواقعية التي حققت نجاحاتٍ تطبيقيةً كبيرةً كثيرةً هامةً، والواقع أن هذه الأدوات الجديدة هي التي تتصدر مشهد الذكاء الحاسوبي الآن في مرحلته المعاصرة.

- بفضل التنامي المتسارع أُسِّيًّا - كما يعبر عنه «قانون مُوور» [2] - في القدرات الحاسوبية التي تتوافــر عليهـــا الحواسيب الرقميـــة، وبفضل بِنَى «الحوسبة على التوازي Parallel Processing» وبِنَى «الحوسبة المـــوزَّعة Distributed Processing» عِلاوةً على ذلك، فقد صار من اليسير في المرحلة المعاصرة من مسيرة «الذكاء الاصطناعي» تنفيذُ خُوارِزماتٍ تقوم بمعالَجاتٍ حسابيةٍ بالغة التعقيد على أحجامٍ ضخمةٍ من البيانات - مما كان متعذِّرًا في مراحله السابقة - بل صارت المنصات الحاسوبية الفائقة في السنوات الأخيرة قادرةً على التهام وهضم «أحجامٍ عملاقةٍ من البيانات».[3] ولذلك تَرْجُحُ الآن بقوةٍ كِفَّةُ «الذكاء القائم على المعرفة Knowledge-oriented Intelligence» - فيما تتمثل المعرفة عن الظاهرةِ محلِّ الدراسةِ بكثافةٍ عاليةٍ من دفق البيانات عن تلك الظاهرة - على كِفَّةِ «الذكاء المقتصر على براعة التحليل والتركيب» كما كان الحال في المراحل السابقة.

١- راجع القسم الفرعي السابق ١,٧ من هذا الفصل؛ وبالتحديد الإطارَ الواردَ بعد الشكل رقم ٤.

٢- الذي سبقت الإشارة إليه في نهاية القسم الفرعي ١,٤ ثم بتفصيل أكبر في نهاية القسم الفرعي ١.٧ من هذا الفصل.

٣- يبلغ معامل تفاوت القدرات في هذا الصدد بين عِقْدِ ستينيات القرن العشرين الميلادي وعام ٢٠١٦م - على سبيل المثال - ما بين «واحدٍ أمامَه أحد عشر صفرًا» إلى «واحدٍ أمامَه اثنا عشر صفرًا».

- وفرت السِّمَاتُ الثلاث السابقة بيئةَ الذكاء الاصطناعي المطلوبة لتفعيل «التعلُّم الحاسوبي machine learning» وتحويله إلى النجم الألمع في العِقْدين السابقين على مسرح الذكاء الحاسوبي مما مَثَّل انعطافةً تاريخيةً ليس فقط في مسيرة تطوير الأنظمة الذكية بل في مسيرة الحوسبة بكاملها؛ ففيها جرى حل الكثير من - أو ربما مُعظَم - ما يمكن مقارَبَتُه تحليليًّا من «مسائلَ مُهَيْكَلةٍ structured problems» ذاتِ بالٍ في المراحل السابقة لمسيرة الذكاء الاصطناعي - وربما كذلك في السنوات الأولى من مرحلته المعاصرة - فإن مُعظَمَ المسائل الهامة الذي بقي إيجادُ حلول حاسوبية ناجعة لها يُشَكِّلُ تحديًّا كبيرًا هي «مسائل غير مهيكلة unstructured problems» لم يَبْقَ من سلاحٍ ماضٍ للتعامل معها سوى «التعلُّم الحاسوبي».[1]

ولأن معظم المساحات الهامة في حقل المعالَجَةِ الحاسوبية الآلية لِلُّغَاتِ الطبيعيةِ - ومن ضمنها «اللُّغَةُ العربيةُ» بطبيعة الحال - تنتمي إلى طائفة «المسائل غير المهيكَلةِ»، فسوف تكون لنا وقفةٌ أخرى مع «التعلُّم الحاسوبي» في القسم الثاني من هذا الفصل؛ حيث نبسِّط الفلسفةَ التي بُنِيَ عليها، وتحت أية شروط يكون هو الخيار الأمثل، وفي أية ظروفٍ ينبغي تجنُّبُهُ.

- في المراحل السابقة كانت النافذة التي يطل من خلالها الذكاء الاصطناعي على مستخدِميه ويطلون منها عليه هي برمجياتٌ يكاد يقتصر كلٌّ منها على تنفيذ خُوارِزمٍ - أو أكثر - من خوارزمات الذكاء الاصطناعي بينما تبقى العناصر البرامجيةُ الأخرى عناصرَ مساعدةً على الهامش، بينما تتميز الحقبة المعاصرة أن تجليات «الذكاء الاصطناعي» تكون من خلال توظيف خوارزماته كعنصرٍ مُدْمَجٍ مع عناصر أخرى عديدة - مثل: «واجهات الاستخدام التفاعلية»، و«الرسوميات»، و«تداول البيانات عبر الشبكات الرقمية»، و«إدارة قواعد البيانات» و«التواصل بين المستخدمين» ... إلخ - تتكامل جميعًا في تطبيقاتٍ رقمية - لا يُشترَط أن ترفع لافتة «الذكاء الاصطناعي» - من أجل تحقيق ذكاءٍ

١- راجع الإطار الأخير في القسم الفرعي السابق ١,٧ لتعريف «المسائل المهيكَلة» و«المسائل غير المهيكَلة» و«التعلُّم الحاسوبي».

حاسوبيٍّ يحس به ويستفيد منه مستخدم تلك التطبيقات. ونشير كالعادة إلى أن برمجيات معالجة اللغات الطبيعية الحديثة ليست استثناءً من ذلك.

- من حيث الصورة الذهنية في الوعي العام أثناء الحقبة المعاصرة؛ فقد خَفُتَ كثيرًا ذِكْرُ مصطلح «الذكاء الاصطناعي» وبَهُتَ أَلَقُهُ سواء في الإعلام واسع الانتشار أو حتى الأوساط الهندسية والأكاديمية، ومقابل ذلك شاع ذِكْرُ الأساليب الناجحة - أو الواعدة - لمقارَبةِ المسائل التي تمثل تحديًا حاسوبيًّا كبيرًا؛ مثل: «نماذج مارْكُوفْ المَخْفِيَّة (Hidden Markov Models (HMMs»، «الشبكات العصبية المحوسَبة (Artificial Neural Networks (ANNs»، «تعلُّم بايِيس الاحتمالي Bayesian Learning»، ... إلخ. استمر ذلك الحال حتى سنوات قليلة خلت حين أخذ الإعلام الإخباريُّ والاقتصاديُّ وحتى الدِّراميُّ واسعُ الانتشار يعيد نَصْبَ رايات «الذكاء الاصطناعي» عاليًا على مِنَصَّاتِهِ ورَفْعَ الصوت بالترويج للمبالَغات في مفتوح آفاقه وخارق قدراته ... إلخ مثلما شهدت المراحل السابقة من هذه المسيرة، ولذلك يجب أن تخضع أيةُ أخبار إعلامية حديثة عن الذكاء الحاسوبي لمزيدٍ من التدقيق والفحص قبل اعتماد صحتها وصدقيتها.

وفي ظل هذه السمات الست العامة المميِّزةِ لحقل الذكاء الاصطناعي في مرحلته الأخيرة المعاصرة فلا بُدَّ من إبرازِ قاعدةٍ ذهبيةٍ ثالثةٍ يجب على من يتصدى في زمننا هذا لمشروعات أبحاث حوسبة اللغات الطبيعية - ومن بينها «العربية» - أو تطوير تقنيات معالجتها أن يضعها نصب عينيه:

قاعدة رقم ٣
إن أي مشروع بحثي أو تطوير تقني جادّ في حقل «المعالجَة الآلية لِلُّغاتِ الطبيعية NLP» هو عملٌ (١) طبيعتُه هندسيةٌ ويجب أن يدار بآليات «الإدارة الاحترافية للمشروعات» خصوصًا مع ضخامتها (٢) وأدواتُهُ هي خُوارِزْماتٌ مبنيةٌ على الرياضيات المتقدمة (٣) وموضوعُهُ ومادتُهُ هي اللغة الطبيعية ومن ضمنها «العربية». وأي عملٍ ذو بالٍ في هذا الحقل يختل فيه واحد أو أكثر من هذه الأركان الثلاثة - من حيث التخطيط أو التجهيز أو الأفراد أو التنفيذ - فإن مصيره المحتَّم إلى الإخفاق.

ماذا تحقق بالفعل إذن حتى الآن في خضم الذكاء الاصطناعي - أو الذكاء الحاسوبي؟ وما الذي يصعب - أو ربما يستحيل - أن يتحقق في المستقبل المنظور؟ هذا سؤال جِدُّ كبير، سنحاول قبل الانتقال إلى القسم الثاني من هذا الفصل أن نرسم ملامح عامةً لإجابته ثم نعود في القسم الخامس من هذا الفصل للإجابة عنه مرةً أخرى بتفصيل أكبر في سياق المعالجة الآلية لِلُّغات الطبيعية.

كان النجاح الأكبر الذي تحقق حتى الآن على مستوى معالجة كميات كبيرة - وأحيانًا هائلة - من الرموز من أجل:

- اكتشاف «الأنماط patterns» وأنساق ارتباطاتها؛ وهو ما يُعْرَف باسم «التنقيب في البيانات Data Mining».
- التصنيف الآلي لهذه الرموز في فئاتٍ معرَّفةٍ سلفًا - وهو ما يعرف باسم «التنسيب classification» - أو مكتشَفةٍ تلقائيًّا من سَلَّة البيانات - وهو ما يعرَف باسم «التصنيف categorization».
- تعميم الربط بين فئتين أو أكثر من الرموز بعضها يمثل مدخَلاتٍ لظاهرةٍ أو عمليةٍ ما وبعضها الآخر يمثل مخرَجاتٍ لها بحيث يمكن بعد ذلك التعميم توليدُ المخرَجاتِ المناظرةِ لأية مدخَلاتٍ وهو ما يُعْرَفُ باسم «التعلم الحاسوبي Machine Learning».
- تعميم الربط بين فئتين أو أكثر من الرموز بعضها يمثل مدخَلاتٍ لظاهرةٍ أو عمليةٍ ما وبعضها الآخر يمثل مخرجاتٍ لها بحيث يمكن بعد ذلك التعميم توليدُ/ اختيارُ أرجح المخرَجات احتماليًّا المناظرةِ لأية مدخَلاتٍ وهو يمكن أن نطلق عليه تسمية «التعلّم الاحتمالي الشرطي Conditional Probabilistic Learning». وتلعب هذه العملياتُ الحداثيةُ [1] لمعالجة الرموز دور البطولة في توفير عنصر الذكاء الحاسوبي في صناعة البرمجيات المعاصرة.

١ - إلى جانب عمليات الذكاء الاصطناعي الكلاسيكية لمعالجة الرموز مثل: التحليل عبر المنطق الرياضي، ومختلَف أساليب البحث الشجريّ والشبكيّ، ... وغيرها مما أسلفنا الإشارة إليه في القسمين الفرعيين ١.٦ و ١.٧ من هذا الفصل.

أما على مستوى «ربط هذه الرموز بالعالم الواقعي»؛ أو بعبارةٍ أخرى «إكساب الرموز مدلولاتٍ في الواقع الحقيقي»، فما زال تحقيقه الآن - وربما لعقودٍ قادمةٍ - بعيد المنال بسبب تعذُّر بناء أونطولوجيا كاملة للعالم الواقعي [1]، وبسبب تعذُّر التمثيل الرياضي وحوسبة «المعرفة الفطرية commonsense knowledge» وتعذُّر حوسبة «التحليل الفِطْريّ commonsense reasoning» تبعًا لذلك، وهي المعضلات التي تناولناها في القسم الفرعي ١,٧ من هذا الفصل [2]. ويُستثنَى من ذلك التعذُّرِ التطبيقاتُ التي تقتصر على عالَمٍ/ بيئةٍ محدودةٍ للغاية؛ مثل «الألعاب الحاسوبية»، و «أنظمة المحاكاة»، وبرمجيات التحكم في الروبوتات المتخصصة ... ففي تلك الحالات نضجت الآن بالفعل الخبرات العملية لبناء الأونطولوجيا التي تُكسِبُ الرموز مدلولاتٍ واقعيةٍ في عالَمها المحدود.

ويبقى تحقيق ذروة الحلم بإكساب واكتساب البرمجيات والآلات التي تنفذها «الذاتية» و «الوعي» - وهو ما يحلو للإعلام الشعبي ودراما الخيال العلمي التعبير عنها بكابوس الآلات المستقلة المتمردة التي تحكم العالم وتستعبد البشر - أقربَ إلى المستحيل استحالةً مبدئيةً بسبب اصطدامهما بسقف الرياضيات المتمثل في «مبرهَنَتَيْ عدم اكتمال الرياضيات Incompleteness theorems» اللتين سبقت الإشارة إليهما في نهاية القسم الفرعي ١,٣ وفي بداية القسم الفرعي ١,٦ من هذا الفصل. ويرجع السبب في ذلك بأوجز وأبسط ما يمكن إلى أن بذرة «عدم الاكتمال» الذي يحدد سقف المعرفة الرياضية تكمن في «استدعاء أية عمليةٍ رياضيةٍ لذاتها» أو «إحالتها على نفسها» وهو ما يصطَلح على تسميته بـ «العَوْديَّة recursion» أو «الإحالة الذاتية self reference» وهو عين المبدأ الضروري للذاتية والوعي حيث إن منشأهما هو معرفة الكائن لنفسه والإشارة إلى ذاتِه. ولذلك فإن أية محاولة ولو أولية لغرس ذاتيةٍ أو وعيٍ في الآلات سوف تضعنا على مسارٍ يؤدي إلى الخوض في مساحاتٍ تعجز أدواتنا الرياضية - وبالتالي نماذجنا الحاسوبية - عن استكشافها ناهيك عن السيطرة عليها. وسوف نعود لتلك المسألة في القسم الخامس من هذا الفصل حين نحاول أن نتلمس سقف المعالجة الآلية لِلُّغات الطبيعية إن كان يحدُّها سقف.

١- وفيما يختص بالمعالجة الآلية للغات الطبيعية فإن بناء أونطولوجيا كاملة للعالم تكسب الوحدات الرمزية للغة مدلولاتها تمثل تحديًا أكبر من ذلك بكثير؛ فالأونطولوجيا المنشودة عندئذٍ ليس عليها فقط استيعاب العالم الحقيقي المحسوس هائل الضخامة بل إن عليها أيضًا استيعاب عوالم المجاز والخيال وهي فضاءات غير محدودة بطبيعة تعريفها!

٢- راجع مناقشة «مفارقة مُورافِيك» قبل الشكل رقم ٤ مباشَرةً، وكذلك الهامشَ الملحَق بهذا الشكل ذاتِه.

٢. مَدْرَستان في معالَجَةِ المسائل في إطارِ الذَّكاءِ الاصطناعيّ.

في الفضاء الشاسع للمسائل - «غير التافهة non-trivial» - المطلوب التصدي لمعالجَتها حاسوبيًّا يمكننا تصنيف الفئات التالية من حالاتها:

(١) «مسائلُ مُهَيْكَلة structured problems»: وهي مسائلُ نعرف عناصرَها ونستطيع أن نفككها إلى مكوِّناتٍ أبسط، وكذلك نعرفُ آلياتِ تفاعُلِ عناصر المسألة مع بعضها البعض، وبالتالي نستطيع أن نضع لها «نُمُوذجًا رياضيًّا صريحًا explicit mathematical model» [١]. وتنقسم «المسائلُ المُهَيْكَلةُ» بدورها إلى فئتين فرعيتين:

(١←١) «مسائلُ مهيكَلةٌ تُوجَد - من حيث المبدأ - حلولٌ لنماذجها الرياضية structured problems with solvable models». وهذه أيضًا تنشعب إلى شعبتين:

(١←١←١) مسائلُ مهيكَلةٌ توجد حلولٌ لنماذجها الرياضية الصريحة ويمكن تنفيذ بعض هذه الحلول حاسوبيًّا بواسطة «خوارزِمات مفيدةٍ» [٢].

(١←١←٢) مسائلُ مهيكَلةٌ توجد حلولٌ لنماذجها الرياضية الصريحة ولا تُعْرَفُ خُوارِزماتٌ مفيدةٌ حاسوبيًّا لتنفيذها عَمَليًّا.

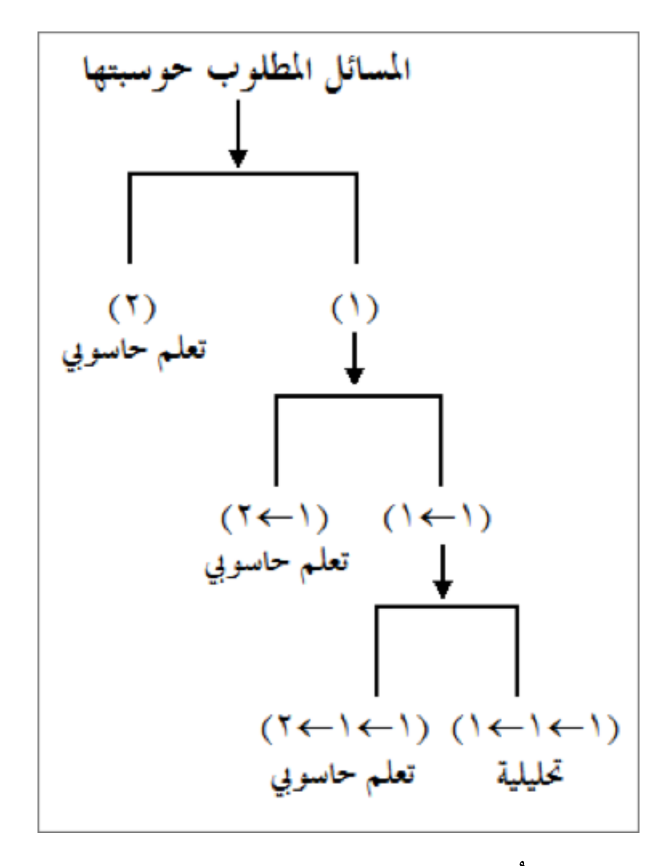
(١←٢) «مسائلُ مهيكَلةٌ ذاتُ نماذجَ رياضيةٍ عَصِيَّةٍ على التعامل معها وحَلِّها نَظَرِيًّا structured problems with unsolvable models».

(٢) «مسائلُ غيرُ مُهَيْكَلة unstructured problems»: وهي مسائلُ يتعذر تفكيكُها إلى مكوِّناتٍ أبسط، ولا تُعْرَفُ آليَّاتُ عملها، وأفضل ما يمكن توفيرُه لوصف مثل تلك المسائل هو عَيِّناتٌ غزيرةٌ من مُدْخَلاتِها بالتوازي مع المخرَجات المناظرة لها.[٣]

١- النموذج الرياضي «الصريح» لنِظام ما هو وَصْفٌ - تقريبيٌّ مبسَّطٌ له في العادة - له عبر نَسَقٍ مترابطٍ من التعريفات والعمليات والعلاقات الرياضية المباشِرة؛ الجبرية و/ أو التفاضلية/ التكاملية و/ أو الشَّرْطية و/ أو المنطقية و/ أو مناولة المصفوفات ... إلخ، بحيث يمكن لهذا الوصف الرياضي إذا أُعْطِيَ قِيَمَ مُدْخَلاتٍ ما أن يُنْتِجَ المخرجَات المناظرة لها باتفاقٍ مقبولٍ مع سلوك النظام الحقيقي. وسوف نتناول مجدَّدًا في القسم التالي من هذا الفصل العلاقةَ بين الظاهرة ونموذجها الرياضي والفوارق بينهما.

٢- راجع تعريف «الخُوارِزم المفيد» الذي عرضناه مع تعريف «الذكاء الاصطناعي» قرب نهاية القسم الفرعي ١,٥ من هذا الفصل.

٣- راجع الإطار الأخير في القسم الفرعي ١,٧ من هذا الفصل.



الشكل ٦: مقارَبةُ فئات المسائلِ المختلِفة حسب تصنيف تعقيدها الرياضي.

ويقترن بهذا التصنيف السابق مقاربتان للتعامل مع المسائل المطلوب حوسبتها:

الأولى: هي «المقارَبةُ التحليليةُ analytic approach» للتعامل مع مسألةٍ ما؛ وهي تنفيذُ حَلٍّ لنموذج رياضي صَريح تنفيذًا حاسوبيًّا. و«النموذج الرياضي الصريح» لنظامٍ ما هو ببساطةٍ وَصْفٌ للنظام عبر نَسَقٍ مترابطٍ من التعريفات والعمليات والعلاقات الرياضية المباشرة، بحيث يمكن لهذا الوصف الرياضي إذا أُعْطِيَ قِيَمَ مُدْخَلاتٍ ما أن يحسبَ المخرجَاتِ المناظرةَ لها باتفاقٍ مقبولٍ مع سلوك النظام الحقيقي.

الثانية: هي مقارَبةٌ «تُعْرَفُ بِاسم «التعلُّم التلقائي» أو «التعلُّم الحاسوبي machine-learning» وهي تنظر إلى النظام المطلوب حوسبته كصندوقٍ مصمَتٍ لا يُعْرَفُ ما بداخله؛ حيث كل ما يُسْمَحُ به هو تغذية النظام بمُدْخَلاتٍ والحصول منه على

المُخْرَجاتِ المناظِرةِ لكلٍّ منها. وركيزة هذه المقارَبة هي تحصيلُ عينةٍ كبيرةٍ (بما يكفي) من المُدْخَلاتِ والمُخْرَجات الموازية لها، ثُمَّ تشغيلُ أحد الآليات الرياضية على هذه العينة الكبيرة من أجل استنباط السلوك العام للنظام بين مُدْخَلاته ومُخْرَجاته.

ويصور (الشكل ٦) عاليَه المقارَبةَ الممكنةَ لكل فئةٍ من فئات المسائل حسب التصنيف السابق وفق طبيعة معرفتنا ببِنْيَتِها الرياضية، وفيما يلي شروط صلاحية وظروف تطبيق كُلٍّ من المقارَبَتَيْن:

- يُمْكِنُ حوسبةُ مسألةٍ ما عَبْرَ «مقاربةٍ تحليليةٍ» فقط عندما تتوافر فيها ما جميع الشروط التالية:
  - معرفة عناصرِ المسألة وإمكانية تفكيكها إلى مكوِّناتٍ أبسط،
  - معرفة آلياتِ تفاعُلِ عناصر المسألة مع بعضها البعض،
  - إمكانية وضع نموذج رياضي «صريح» لهذه المسألة،
  - وجود حلول نظرية لهذا النموذج الرياضي الصريح،
  - إمكانية تنفيذ أحد هذه الحلول النظرية حاسوبيًّا بواسطة «خُوارِزمٍ مفيد».
- إذا افتُقِدَ واحدٌ أو أكثرُ من شروط إجراء المقارَبة التحليلية - السابق ذكرُها - فلا يتبقى خيارٌ عندئذٍ إلا مقارَبة المسألة عبر «التعلم الحاسوبي».

ويُنْظَرُ عادةً إلى «المقارَبة التحليلية» على أنها الأسلوبُ التقليديُّ لحوسبة المسائل، وإلى مقارَبة «التعلُّم الحاسوبي/ التلقائي» على أنها هي الممارَسةُ «الأكثر حَدَاثيَّةً» لأن المقارَبةَ الأخيرة تحتاج بالفعل إلى رياضياتٍ أكثرَ تقدُّمًا وتعقيدًا وكذلك إلى قدراتٍ حاسوبيةٍ لتنفيذها أعلى من الأولى ولذلك تأخر تطويرُها زمنيًّا. ولكنَّ المفارَقةَ اللافتةَ هي أن المقارَبةَ الفِطْريةَ في الطبيعة وكذلك في سلوك البشر هي «التعلُّم الحاسوبي/ التلقائي»؛ فالطفل الصغير مثلاً يتعلم اللغة بالمعايشة والمحاكاة دون أن يعرف أو يدرس أولاً نظامها الصوتيَّ ولا الصرفيَّ ولا النحويَّ ... إلخ، والصَّبِيُّ الذي يتعلم حِرْفةً كالحِدادة والسَّبْك» يتعلمها أيضًا من مُعَلِّمِهِ (أو معلِّميهِ) بالمعايشة والمحاكاة دون أن يعرف أو يدرس أولاً علمَ المعادن ولا المعالَجاتِ الحرارية ولا فيزياءَ المواد ... إلخ.

وكمثالٍ على المقارَبةِ التحليلية نذكر المسألةَ (الجُزْئيَّةَ البسيطةَ) الخاصةَ بالتحقق من «كلمة السر password» الرقمية لحساب المستخدِم على أية منصة برامجية حاسوبية؛ فمن الواضح أن عملية التحقق يمكن إجراؤها عن طريق اشتراط المطابَقة الكاملة بين تسلسل الرموز المُدْخَلة وبين تسلسل الرموز في كلمة سر المستخدِم المحفوظة (بعد فك تشفيرها)، وهو ما يتضمن تعريفًا رياضيًّا (للرموز المسموح بها وغير المسموح بها في كلمات السر) وعمليةً رياضيةً (اختبار تساوي أو عدم تساوي رمزين) وعلاقةً رياضيةً (علاقة التساوي)، وهي كلها صريحة مباشرة.

كما نذكر كمثالٍ على مقارَبة التعلم الحاسوبي مسألةَ «التعرُّف على مفرَدات الكلام المنطوق [1] Speech Recognition» من متحدَّثٍ واحدٍ التي سبق عرضُها قُبيل ختام القسم الفرعي ١,٧ من هذا الفصل. وأفضل ما يُعْرَفُ من أساليبَ رياضيةٍ - دون الولوج في تفاصيلها - لإنجاز التعلُّمِ الحاسوبي مع هذه «المعضلة» هي «نماذج ماركوف المَخفيَّة Hidden Markov Models» بالتزاوج مع «النماذج الاحتمالية لتراكيب الفونيمات».

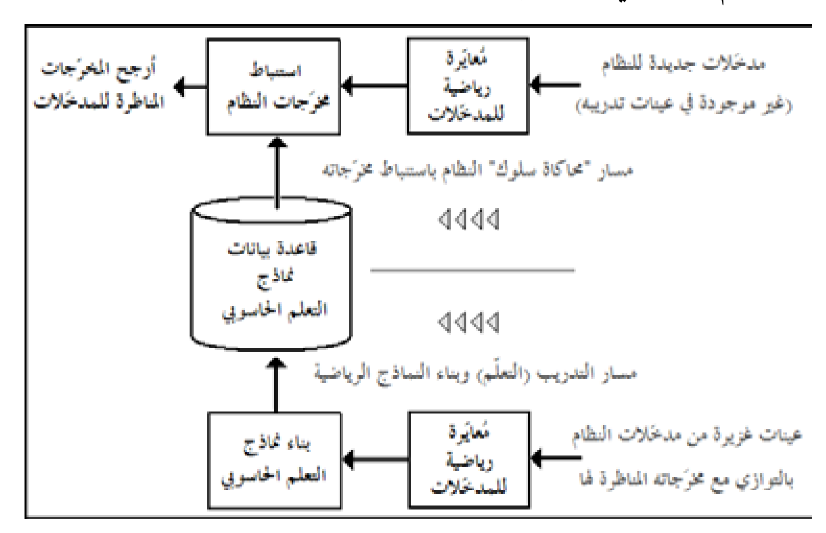
ويبين (الشكل ٧) أدناه المكوِّناتِ العامةَ لمقارَبةِ التعلُّم الحاسوبي؛ حيث تَظْهَرُ ارتباطاتُ هذه المكونات ومساراتُ تدفُّق البيانات فيما بينها. ومن أهم ما ينبغي ملاحظتُه في هذا الشكل وجود مسارَيْن؛ أولهما تُهْضَمُ فيه عيناتٌ وفيرةٌ من مدخَلات النظام/ الظاهرة قَيْدَ الحوسبة والمخرَجات المناظرة لها - بعد تحويل صورتها الخام إلى صورةٍ رياضيةٍ نمطيةٍ/ عياريةٍ - عبر واحدٍ من أساليب التعلُّم الحاسوبي حتى تُحْسَبَ منها معامِلاتُ نماذج التعلُّم الحاسوبي وتخزَّنَ في قاعدة بياناتٍ، ويسمَّى هذا المسار «طَوْر التعلُّم learning phase» أو «طَوْر التدريب training phase». أما ثاني المسارين فيُسَمَّى «طَوْرُ الاستنباط» أو «طَوْرُ التشغيل» وهو الغرض من هذه المقارَبة كلها؛ وفيه تُستنبَطُ مُخرَجاتُ النظام/ الظاهرة المقابلةُ لأية مدخَلاتٍ له ولو لم تكن في عينات طور التدريب. ومع اختلاف التفاصيل الرياضية فإن كلا الطَّوْرَيْن رُكْنان أساسيَّان في مختلَفُ أساليب التعلم الحاسوبي المعروفة مثل «الشبكات العصبية المحوسبة Artificial Neural Networks ANN»، و «آلية المتجه الداعم Support Vector Machine»،

١- بافتراض أن «حصيلة المفرَدات vocabulary» المطلوب التعرف عليها محدودةَ العدد مغلَقة.

والتعلم وفق «مبدأ حالة اللا يقين القصوى Maximum Entropy Principle»، و «تعلُّم بايِيس الاحتمالي Bayesian Learning»، ... إلخ.

وينبغي التنبيهُ على أنه إذا توافرت شروطُ «المقارَبة التحليلية» المذكورة سلفًا في معالَجة مسألةٍ ما فلا ينبغي اللجوء إلى مقارَبة «التعلُّم الحاسوبي» في معالجَتِها؛ حيث إن الأولى هي الأوضح والأضبط تعبيرًا عن طبيعة المسألة وهي في الغالب الأكثر اقتصادًا في الإنجاز المعلوماتي الحاسوبي، أي أنها كالماء لوُضوء الصلاة إذا حضر بَطُلَ التيمم.

بَقِيَ قبل الانتقال إلى القسم الثالث من هذا الفصل أن نشير إلى أن بعض المسائل المركَّبة قد تَصْلُح لمعالجَة جزءٍ منها «المقارَبةُ التحليلية» بينما لا يَصْلُح لمعالجَة بقيتها إلا مقارَبة «التعلُّم الحاسوبي» وعندئذٍ تسمَّى «المقارَبة الهجينة Hybrid approach».



الشكل ٧: مكوِّناتُ مقارَبةِ التعلُّم الحاسوبي، وارتباطاتُها، ومساراتُ تدفُّق البيانات بينها. [١]

٣. تحديات نَمْذَجة اللُّغات الطَّبيعية ومعالَجتها آليًّا.

سوف نُسْقِطُ في هذا القسم بعض الاعتبارات الهامة للمقارَبتَيْن «التحليلية» و «التعلُّم الحاسوبي» على قضية «المعالَجة الآلية اللغات الطبيعية Natural Language Processing (NLP)»، ولكن قبل ذلك نحتاج إلى التوقف مع

١- تشتهر («معامِلات نماذج model parameters» التعلم الحاسوبي) أي «القيم العددية المحدَّدة لحالات هذه النماذج» بين الباحثين والمطوِّرين في هذا الحقل بالاسم المختصَر («نماذج models» التعلم الحاسوبي)؛ لذا لزم التنبيه على عدم الخلط بينهما.

اصطلاحٍ مركزيٍّ تكرر معنا مِرارًا منذ بداية هذا الفصل لتأمُّل معناه وعواقِبه ألا وهو «النموذج الرياضي Mathematical Model» و «النَّمْذَجة Modeling». فمن أجل دراسةِ ظاهرةٍ/ مسألةٍ طبيعيةٍ ما على نحوٍ علميٍّ موضوعيٍّ يسمح بالقياس والمقارَنة والتنبُّؤ والتصميم والتخطيط والتنفيذ فإن أفضل ما لدينا من أدواتٍ هي ترسانةُ الرياضيات بما فيها من حسابٍ ومنطقٍ وتحليلٍ ... إلخ، ومن أجل إعمال هذه الأسلحة والحيل الرياضية في المسألة قَيْد الدراسة فلا بُدَّ أولاً من صِياغةِ وصفٍ رياضيٍّ متماسكٍ لهذه المسألة؛ فيسمَّى هذا الوصف «النُّمُوذَجُ الرياضي» للمسألة كما تُسَمَّى عملية بنائه «النَّمْذَجةُ الرياضية» لها.

وهنا يبرز سؤالٌ محْوريٌّ: هل النموذج الرياضي مُكافِئٌ للمسألة الطبيعيةِ المُنَمْذَجةِ؟ والإجابة قولاً واحدًا هي أنه يستحيل[1] أن يكافئَ نموذجٌ رياضيٌّ المسألةَ التي ينمذجها، وأن أفضلَ نموذج رياضيٍّ يمكن بناؤه للمسألة هو نُموذَجٌ تقريبيٌّ مبسَّطٌ لها. ويبرز (الجدول ٢) أدناه أهم الملامح التي تختلف فيها الظواهر/ المسائل الطبيعية عن أفضل ما يمكن بناؤه من نماذج رياضية لها.

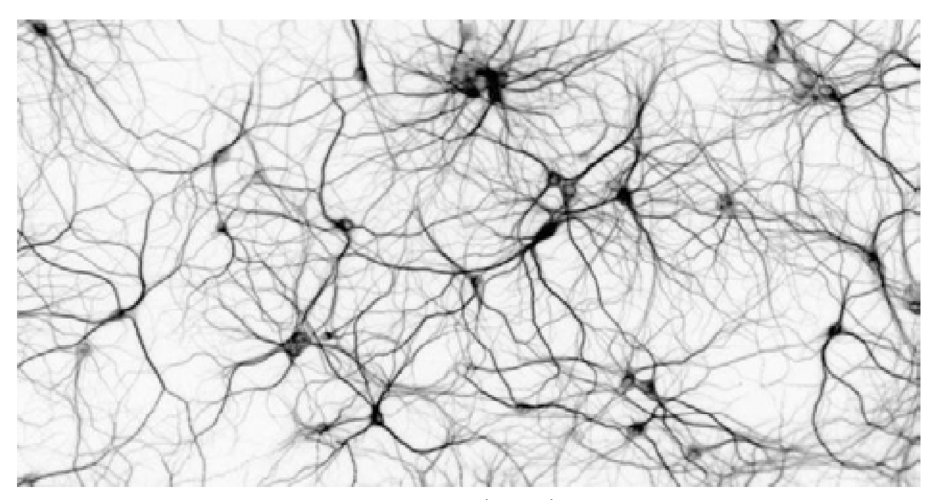
صفات الظاهرة الطبيعية	صفات النموذج الرياضي للظاهرة الطبيعية
الكثرة الهائلة للعناصر والمكوِّنات	محدودية عدد العناصر والمكوِّنات
التعقيد الهائل (شبه الفوضوي) لترتيب العناصر وعلاقا تها	تنظيم وترتيب عناصر النموذج في أنساق بسيطة
الاتصال والاعتماد المتبادَل بين مختلَف العناصر	تقليل الاتصال والاعتماد بين العناصر إلى أدنى حد
لا حدودَ حاسوبيةً تقيد تطور تفاعلات الظاهرة	وجودُ حدودٍ حاسوبيةٍ مقيِّدةٍ لتنفيذ النموذج
لا قيودَ حاسوبيةً على دِقَّةِ تفاعلات الظاهرة	وجود قيود على الدقة الحسابية في تنفيذ النموذج
كل العناصر والتفاعلات مضمَّنة في الظاهرة	بعض العناصر وآليات التفاعل مجهولة
كل العناصر والتفاعلات مشارِكة في الظاهرة	بعض العناصر والتفاعلات مهمَلة في النموذج
الاتصال والاعتماد المتبادَل بين الظاهرة وبيئتها الخارجية	النموذج قائمٌ بذاته معزولٌ عما سواه

الجدول ٢: أهم الفوارق بين الظاهرة الطبيعية وأفضل نموذج رياضي يمكن بناؤه لها.

١- وهي استحالةٌ مبدئيةٌ لأن تلك الفوارق تظل قائمةً في عمق الطبيعة وظواهرها وفي عمق الرياضيات وعلوم الحاسب، راجع القسمين الفرعيين ١,٣ و ١,٤ من هذا الفصل، فلربما تتضاءل هذه الفوارق بعض الشيء مع تحسُّن تقنياتنا لكنها لن تتلاشى.

وكمثالٍ تطبيقيٍّ على هذه الفوارق نشير إلى «الشبكات العصبية» كظاهرةٍ حقيقيةٍ مقابل «الشبكات العصبية المحوسَبة» كنموذج رياضي لها [١]، ونترك لفطنة القارئ تعيينَ الفوارق التي ذكرناها لتونا على هذا المثال بالاستعانة بالأشكال الثلاثة التالية (أرقام ٨ و ٩ و ١٠) التي تصور على الترتيب «الشبكة العصبية الحقيقية»، ثم «التبسيط الأَوَّليّ لها»، ثم «الشبكات العصبية المحوسَبة متعددة الطبقات» وهي الآنَ النموذجُ الرياضيُّ المعياريُّ للظاهرة.

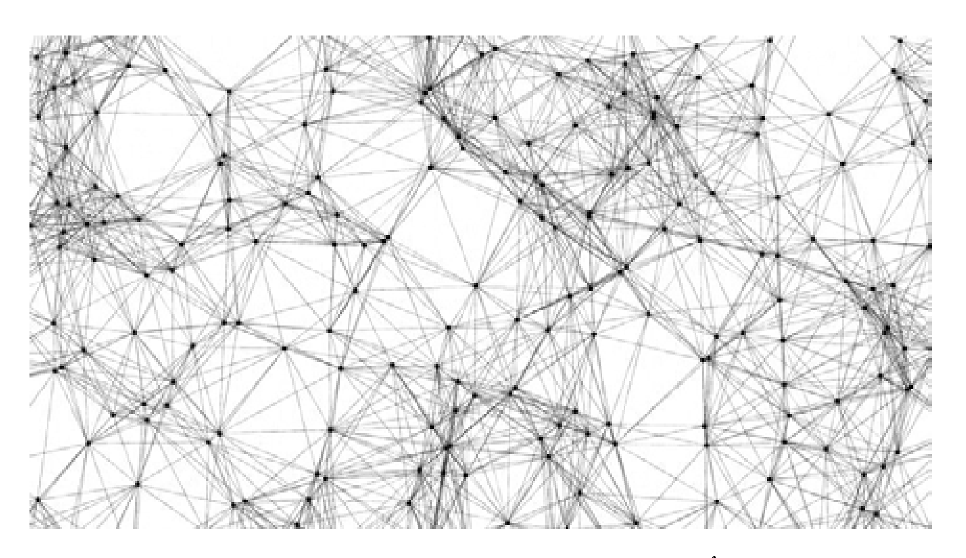
وبسبب تلك الفوارق بين الظاهرة الحقيقية ونموذجها الرياضي، فلا مَفَرَّ من بقاءِ هامشِ خطأٍ غير صِفْريٍّ (ولنُسَمِّهِ **ط**) عند تنفيذ النموذج، ومن الهام للغاية الانتباه عند المفاضلة بين نموذجين رياضيين لنفس الظاهرة/ المسألة أن المقارَنة لا تكون بين دقة الأداء (ولنُسَمِّها **ق**) لكلا النموذجين – وهي مغالَطةٌ شهيرةٌ يلجأ إليها كثيرًا بعض المدلِّسين – بل بين هامشَي الخطأ فيهما؛ وبعبارةٍ أخرى فإن: معامل المفاضلة بين نموذج١ ونموذج٢ يساوي: $ط_{٢} \div ط_{١} = (١-ق_{٢}) \div (١-ق_{١})$ بينما المغالَطة هي أن يُحْسَبَ هذا المعامل بالصيغة $ق_{١} \div ق_{٢}$. [٢]



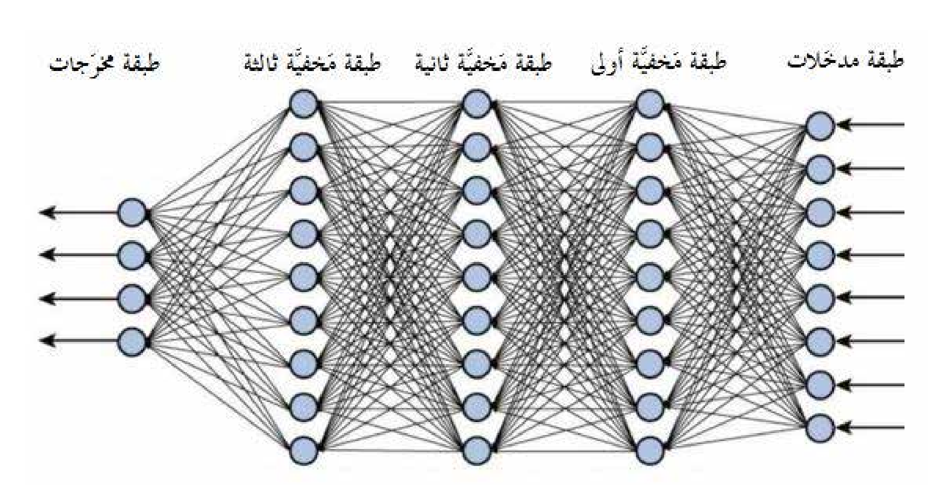
الشكل ٨: لقطةٌ مِجْهريةٌ من شبكةٍ عصبيةٍ حقيقية.

١– وقد تعرضنا للشبكات العصبية وحوسَبتها في الإطار قبل الأخير من القسم الفرعي ١,٧، ثم في القسم الفرعي ١,٨ من هذا الفصل.

٢– مثال: مع نموذج١ ذي دقة أداء ٩٠٪ مقابل نموذج٢ ذي دقة أداء ٩٥٪، فإن معامل أفضلية الثاني على الأول يساوي (١ – ٩٠٪) ÷ (١ – ٩٥٪) = ٢ ولا يساوي أبدًا ٩٥٪ ÷ ٩٠٪ = ١٩÷١٨، والفارق بين التقديرين كبير جدًّا.



الشكل ٩: تبسيطٌ أوليٌّ للشبكات العصبية يحافظ على القنوات الاتصالية بين خلاياها.



الشكل ١٠: النموذج المعياري للشبكات العصبية المحوسَبة متعددة الطبقات.

ورجوعًا إلى قضية اللغات الطبيعية – وهي الظاهرةُ الطبيعيةُ شديدةُ التعقيد – فإن التحديات التي تواجه نمذَجتَها حاسوبيًّا بكُلِّيَّتِها كبيرةٌ للغاية نعيد إجمالَ أهمِّها فيما يلي:

أولاً: تحدي الربط بين اللغة من حيث هي رموزٌ (أصوات، مُفْرَدات، تراكيب ... إلخ) وبين دِلالاتها في العالم الواقعي؛ حيث إن «المعرفة الفِطرية commonsense knowledge»[1] بالعالم الحقيقي – وهي التي يشترك فيها كل (أو معظَم) البشر – هائلةُ الضخامة لدرجةٍ تجعل من تمثيلها رياضيًّا في قاعدة معرفية (تُعْرف باسم «أُونْطولوجْيَا») أمرًا بعيد المنال، ناهيكَ عن المعارف المتخصصة لفئة معيَّنة من الناس (أونطولوجيا عُمَّال المناجِم، أونطولوجيا أطباء الأسنان ... إلخ).

ثانياً: لأن الخيال جزء أساسي في التفكير البشري فإن التفكير فيه والتعبير عنه ينعكس كذلك في اللغة الطبيعية، ولذلك فإن التحدِّيَ المُعْجِزَ لا يقتصر فقط على بناء أونطولوجيا للعالم الحقيقي بل يمتد إلى الحاجة إلى بناء أونطولوجيا للخيال (وهو مجال مفتوح غير محكوم)!

ثالثًا: لا يُضَمِّنُ «مُنتِجُ اللغة» (المتحدِّثُ أو الكاتبُ) عند استخدامه اللغةَ أداةً اتصاليةً أو قناةً لنقل محتوًى معرفيٍّ على كُلِّ المعلومات التي يريد توصيلَها للمتلقِّي (السامع أو القارئ) في رسالته، فلو فعل لصارت العمليةُ ثقيلةَ العبء لدرجة تقارب الاستحالة، ولكنه يفترض أن السامع يشترك معه في معرفة سياق رسالته[2] ولذلك فإنه يُضَمِّنُ في رسالته فقط المعلوماتِ المضافةَ على سياقها، فتبقى المادة اللغوية (نصًّا أو كلامًا منطوقًا) خُلْوًا من السياق الذي تجري فيه الرسالة وتكون بذلك المعلومات المتاحة للآلة التي تحاول معالجة اللغة الطبيعية ناقصةً نقصًا فادحًا.

رابعًا: يفترض كذلك مُنتِجُ اللغةِ في متلقِّيها أنه قادرٌ على استكمالِ كثيرٍ من المحذوفات في رسالته ليس فقط لامتلاكه «المعرفة الفطرية»، بل كذلك لامتلاكه القدرة على «التحليل الفطري commonsense reasoning»[3] وهذا تحدٍّ آخَرُ مُعْجِزٌ.

١– سبقت الإشارة إلى تلك المسألة الأونطولوجية في هذا الفصل في القسم الفرعي ٦, ١ – بعد الشكل رقم ٣ – وكذلك قرب ختام القسم الفرعي ٨, ١.

٢– متكلِّمٌ بتذمُّر: «غريبٌ أمر هذا الطقس!». السياق الغائب هو: أن الكلامَ في وقت الصيف في مكانٍ يكون عادةً ما يكون جافًّا صيفًا بينما عاصفةٌ ماطرةٌ تمر بالمكان.

٣– راجع التعليقَ الهامشيَّ المقترنَ بالشكل رقم ٤ في القسم الفرعي ٧, ١ من هذا الفصل حيث فصَّلنا مُعضِلةَ التحليل الفِطْريَّ.

خامسًا: لا يخدم «المجَازُ metaphor» في اللغة فقط أغراضَها الشعريةَ والتصويرَ الأدبي، بل إنه في الحقيقة يتغلغلُ تغلغلاً لا ينفصم عن صُلْبِ أي غرضٍ من أغراض استخدامها [١]، والتحدي الذي يمثله المجاز هو أنه يزرع في الرسالة اللغوية عاملًا من «الالتباس المقصود « لا سبيل لإزالته آليًّا حتى ولو توفرت لدينا - بمعجزةٍ - أونطولوجيا للخيال البشري.

سادسًا: كل ما سبق من تحدياتٍ ليست «استاتيكيةً» (ثابتةً في الزمن) بل «ديناميكيةً» (متحرِّكةً) دائمًا فيه، وتعتمد هذه الحركة في الزمن على تفاعلاتٍ اجتماعيةٍ ثقافيةٍ غايةً في التعقيد والتشابك، ولذلك فحتى لو تغلبنا - من باب الافتراض – على ما سبق من تحديات فإن علينا تعديلها باستمرار لتساير هذه التغيراتِ الدائبة.

سابعًا: ونأتي إلى مسك ختام هذه التحديات وأعسرها جميعًا ألا وهو تحقيق عامِلَيْ «الوعي» و «الذاتية» في اللغة والذي (يكاد) يستحيل مبدئيًّا تحقيقُهما لاصطدامهما بسقف الرياضيات؛ ويُرْجَى الرجوع إلى ختام القسم الفرعي ١,٨ من هذا الفصل حيث سبق أن بَيَّنَّا فيه سبب تلك الاستحالة. وتعني تلك الاستحالةُ أن أية آلة/ برمجية حاسوبية لمعالجة اللغات ستظل مجرَّدَ آلةٍ جامدةٍ لا تعي وجودَها، ولا ذاتيةَ لها، تنفِّذ فقط ما يُمْلَى عليها وما تُمْلأُ به من تعليماتٍ مهما كانت.

النتيجةُ الواضحةُ إذن هي أن أَيَّةَ نماذجَ رياضيةٍ صريحةٍ تحاول استيعابَ «كامل» الظاهرة اللغوية الطبيعية عليها أن تتغلبَ على كل الأهوال السابق ذكرُها وهي في مقام المستحيلات أو تكاد. ولذلك فإن نمذجة الظاهرة اللغوية الطبيعية بأكملِها عبر «المقاربة التحليلية» لربما كانت أكثر مهام النمذجة التحليلية عُسْرًا على الإطلاق، ولِمَ لا والإنسان هو أعْقَدُ الكائنات على الإطلاق، والظاهرةُ المطلوب استيعابُها هي لغتُهُ التي ينعكس فيها جوهر فكره ويتردد فيها صدى وجوده.

١- « ... وأضاف أنه يتطلع إلى تحطيم الأرقام القياسية للأرباح التي حققتها الشركة في العام المالي المنصرم ...» هذه جُذَاذةٌ من تقريرٍ ماليٍّ جادٌّ عن أعمال إحدى المؤسسات التجارية، وتحتوي الجذاذة عدة استخدامات مجازية؛ منها مثلاً «يتطلع» حيث جَسَّدَ المتكلم «تحطيم الأرقام القياسية» وكأنه شيءٌ ماديٌّ يُرَى، وكذلك «تحطيم» حيث جَسَّدَ المتكلم أيضًا «الأرقام القياسية» وكأنها شيءٌ ماديٌّ قابلٌ للتحطيم، ... إلخ.

لا مَفَرَّ إذن من مقاربة «التعلم الحاسوبي» إن كنا نريد أن ننجز أي تقدم على صعيد المعالَجَة الآلية «الكاملة» للغات الطبيعية، أليس كذلك؟ ولكن ماذا لو كنا فقط نطمح إلى معالَجَةٍ آليةٍ «جُزْئيَّةٍ» لِلُّغات الطبيعية؛ مثل «الصَّرف morphology» أو «النحو syntax»؟ هل تُجْدِي عندئذٍ نفعًا «المقارَبةُ التحليلية» وَحْدَها؟

يمكن بالطبع بناءُ نموذج رياضي تحليلي لبنية الكلمة وصفاتها في أية لغة طبيعية في كلا الاتجاهين البنائي والتفكيكي ومن ثم تنفيذه حاسوبيًّا، ولكن المشكلة التي سنكتشفها سريعًا هي تعدد الحلول التي ينتجها مثل ذلك النموذج؛ وهو ما يعرف بـ «الالتباس ambiguity» والذي لا يمكن إزالتُهُ - أي اختيار الحل الصحيح من بينها - إلا بالتغلُّبِ على التحديات التي أسلفنا ذكرَها.[١] وكذلك يمكن (بصعوبة) بناء نموذج رياضي تحليلي يستوعب التراكيب النحوية المسموح بها في أية لغة طبيعية في كلا الاتجاهين البنائي والتفكيكي ومن ثم تنفيذه حاسوبيًّا، ولكن القاعدة الأصولية الشهيرة لدى النُحَاة «الإعرابُ فَرْعُ المعنى» تكفينا مؤونة شرح ما سيحدث حينئذٍ من التباسٍ عنيفٍ[٢] في غياب معالجة التحديات السابق ذكرها آنفًا.

وفي الحقيقة فإنه يندر العثورُ على مسائلَ لمعالَجَةِ اللغات الطبيعية -ولو جُزْئيًّا - يمكن نمذجتُها دون التباسٍ بمقارَبةٍ تحليليةٍ صِرْفة، فهل التزامُ مقارَبةِ «التعلُّم الحاسوبي» وإهمالُ «المقارَبة التحليلية» هو القَدَرُ الذي لا فكاكَ منه دائمًا؟ الشِّقُّ الأولُ من الإجابة هو: نعم، مقارَبةُ «التعلُّم الحاسوبيُّ التلقائيُّ machine learning» ضروريةٌ

١- عندما صادفتك عزيزي القارئ هذه الفقرة «يمكن بالطبع بناءُ نموذج رياضي تحليلي لبنية الكلمة وصفاتها في أية لغة...» فلا بد أنك قد فَسَّرْتَ أن كلمةَ «لِبِنْيَةِ» هي (حرف جر «لِـ» + اسم مجرور على وزن «فِعْلَة» هو بِنْيَة من مادة «ب ن ي») لأنك تفهم النظام الصرفي العربيَّ ومنظومَته النحويَّة وتعرف سياقَ الكلامِ ودِلالَتُه وتستنتج مَغْزَى كاتبه ... إلخ. ولكن أي نموذج تحليلي يغطي النظام الصرفي العربي فقط سوف ينتهي إلى حلولٍ متعددة: («لَبَنِيَّة» = اسم على وزن «فَعَل» هو «لَبَن» من مادة «ل ب ن» + ياء النسب + تاء التأنيث) أو (لِبُنَيَّة: حرف جر «لِـ» + اسم مجرور على وزن «فُعَيْل» المحوَّر بسبب الإعلال إلى «فُعَيّ» هو «بُنَيّ» من مادة «ب ن و» + تاء التأنيث) أو (لِبُنِّيَّة: حرف جر «لِـ» + اسم مجرور على وزن «فُعْل» مضعَّف العين واللام هو «بُنّ» + ياء النسب + تاء التأنيث) ... إلخ.

٢- الواقع أن الالتباس الحادث حينئذٍ ذو رُتْبةٍ أعنفَ كثيرًا من الالتباس الصرفي الذي سبقت الإشارة إليه والتمثيل له؛ حيث إن البنية النحوية هي بِنْيةٌ شجريةٌ في جوهرها (لأن كل مُكوِّنٌ من مكوِّنات التركيب النحوي قد يكون هو نفسه تركيبًا نحويًّا) ولذلك ينفجر عدد المسارات أُسِّيًّا أثناء إجراء التحليل - راجع مناقشة مشاكل البحث الشجري في بدايات القسم الفرعي ١,٧ من هذا الفصل - وبالتالي يتعذر توليد كل الحلول النحوية الممكنة من أجل الاختيار بينها، بل يتحتم تقليم المسارات المولَّدة أثناء إجراء التحليل أولاً بأول من أجل كبح جماح هذا الانفجار الأُسِّيّ.

للتعامل الحاسوبي «الخالي من الالتباس» مع الغالبية العظمى من مسائل «المعالَجَة الآلية لِلُّغات الطبيعية». أما الشق الثاني من الإجابة فهو أنه لا يمكِنُ إهمالُ «المقارَبة التحليلية analytic approach» فهي أيضًا ضروريةٌ إلى جانب مقارَبة «التعلُّم الحاسوبي»! ومَرَدُّ ذلك إلى سَبَبَيْن؛ أولهُما وأهمُّهما هو أن تصميم وبناء المُدْخَلات إلى آلية التعلُّم الحاسوبي في طَوْرَيْها التدريبيِّ والتشغيليِّ - راجع (الشكل ٧) أعلاه - تتضمن بناء نماذج تحليلية للوحدات/ التراكيب اللسانية الأساسية في الظاهرة اللغوية الطبيعية المطلوب تعلُّمُها حاسوبيًّا. أما ثاني هذين السببين فهو أنه في بعض الأحيان يَحْسُنُ - لأسبابٍ تتصل باقتصاديات صناعة البرمجيات وبكفاءة أدائها - أن يجري حل مسألة المعالَجَة اللغوية الطبيعية حلًّا أوليًّا جزئيًّا عبر «المقارَبة التحليلية» ثم يجري إزالة الالتباس الناتج من هذه المقارَبة عبر مقارَبة «التعلُّم الحاسوبي» وتُعْرَفُ تلك الممارَسة عندئذٍ باسم «المقارَبة الهجينة hybrid approach».[١]

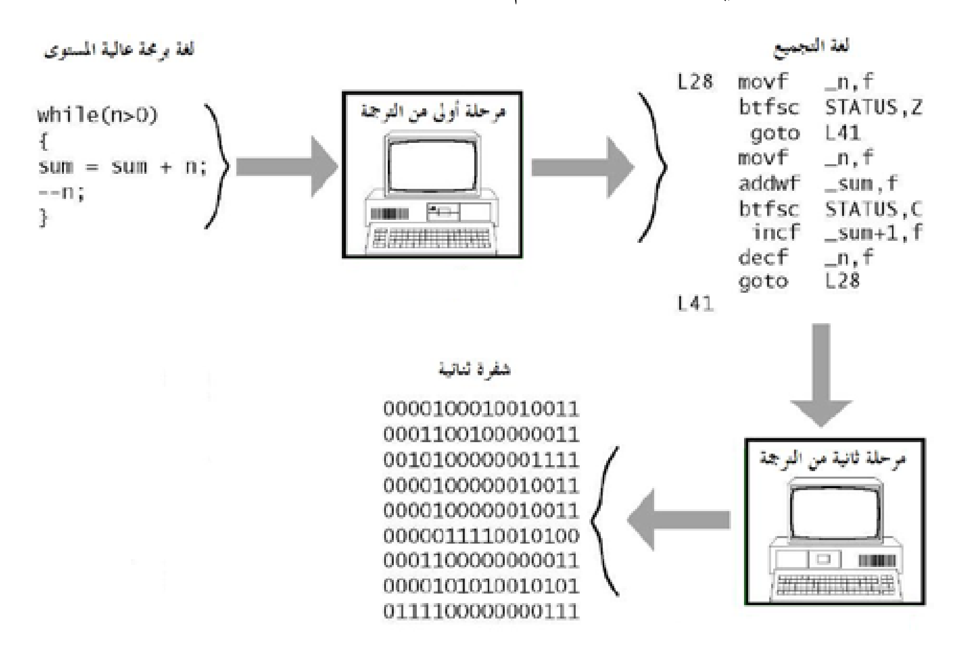
لَزِمَ قبل الانتقال إلى القسم الرابع من هذا الفصل تحريرُ مسألةٍ سببت وما زالت تسبِّبُ كثيرًا من الجدل والخلط، وأدت بكثيرٍ من الباحثين الساعين بإخلاصٍ في طريق «المعالَجَة الآلية للغة العربية» إلى التعثُّر في مسيرتهم البحثية والتطويرية؛ حيث لا يزال يشيع بينهم أن الطريق الواعد لتحقيق بعض الإنجازات الكبرى كالترجمة الآلية من وإلى العربية، وكالتحليل النحوي الآلي لِلُّغة العربية، ... إلخ يقع في دُرُوب ما يشتهر بمدرسة «تشُومِسْكِي» [٢] من «نحو تحويلي» و «نحو توليدي» ... وما إلى ذلك من «مقارَباتٍ تحليلية» أُسِّسَتْ على نظرية «النحو الرياضي Formal Grammars» و «اللغات المحدَّدة رياضيًّا Formal Languages». فما هي حقيقة ذلك وما القصة وراءه؟

الحقيقة هي أن هذه المدرسة بأساليبها قد حققت نصرًا عمليًّا هائلاً بدأ من نهاية خمسينيات القرن العشرين الميلادي ويمتد أثره العميق حتى وقتنا الراهن؛ ألا وهو

١- من الأمثلة التي تصلح معها تلك «المقارَبة الهجينة» التعامل مع مسألة «التحليل الصرفي العربي» التي أشرنا إليها سابقًا؛ فالمقارَبة التحليلية عندئذٍ تستنتج كل الحلول الممكنة على القياس في نظام الصرف العربي، ثم تتكفَّل مقارَبة «التعلُّم الحاسوبي» باختيار أرجح تلك الحلول بمعلومية اقتراناتها الإحصائية بما حولها من مفرداتٍ وتراكيب.

٢- «نَعُوم تشُومِسْكِي» هو العالم الأمريكي البارز في «العلوم الإدراكية»، و«علوم الحاسب» وهو الأب المؤسس لِلِّسانيات المعاصرة، وهو كذلك «فيلسوفٌ تحليليٌّ»، ثم مؤرِّخٌ وناشطٌ سياسيٌّ نافذٌ، وهو أستاذ في «معهد ماساتْشُوسِيتْس للتقنية MIT» ثم في «جامعة أرِيزُونا» بالولايات المتحدة الأمريكية، ويعتبره الكثيرون من المؤرخين أحد أهم الشخصيات الفكرية المئة التي أثرت في المسيرة الحضارية للقرن العشرين الميلادي.

تمكنها من تصميم وتنفيذ مختلَف «لُغَات برمجة الحواسيب عالية المستوى High Level Languages» وهي التي نقلت صناعة البرمجيات والمعلوماتية الحاسوبية إلى مستوى الصناعات الثقيلة الأساسية في الاقتصاد المعاصر. فقبل لغات البرمجة عالية المستوى تلك كان على من يود «إعطاءَ تعليماتٍ» - أي «كتابة برنامج» - لحاسوبٍ رقميٍّ أن يكتبها بـ«الشفرة الثنائية Binary machine code» التي لا تفهم الوحدات الإلكترونية الحسابية والمنطقية سواها، وهو أمر صعبٌ مكلِفٌ وخصوصًا مع المهام البرمجية الكبيرة فضلاً عن ارتفاع احتمالية وقوع الأخطاء أثناء عملية البرمجة. أما لُغاتُ البرمجة عالية المستوى فهي تشبِهُ لغة صياغة الخُوارِزمات بنكهةِ اللغة الإنغْليزية وهو ما يتقنه المتخصصون في الرياضيات وعلوم الحاسب، ومهندسو المعلوماتية والبرمجيات.



الشكل ١١: الترجمة من «لغةٍ برمجيةٍ عاليةِ المستوى» إلى «شفرةٍ ثنائية».

وهنا يأتي الدور «السحري» لتلك المدرسة في القيام بالترجمة الآلية من اللغة البرمجية عالية المستوى إلى الشفرة الثنائية؛ وهو ما يصور (الشكل ١١) عالِيَه مثالاً عليه. أغرى هذا النجاح العظيم الكثيرين من الباحثين إلى محاولة تكراره في ملعب اللغات الطبيعية واعتماد تلك النماذج التي نجحت مع «اللغات عالية المستوى لبرمجة الحواسيب

الرقمية» في حوسبة اللغات الطبيعية توطئةً لمعالجتها آليًّا. وبعد أكثرَ من ثلاثة عقودٍ أثمرت المحاولات الحثيثة في ذلك المضمار عن إجراءاتٍ وأساليبَ شتى لتحليل اللغات الطبيعية أفادت كثيرًا اللِّسانيِّين (البشريين) في تطبيقاتٍ كالنَّقْد، وتحليل الخطاب، وعلم اللغة المقارَن، وتحليل الإعاقات اللغوية ... إلخ، لكنها لم تفلح حتى الآن في تقديم حلولٍ عمليةٍ خاليةٍ من الالتباس للمعالجَة الآلية (الحاسوبية) لِلُّغات الطبيعية.

والسبب الجَلِيُّ في ذلك الإخفاق هو أن «لغات البرمجة عالية المستوى» قد صُمِّمَت ببالغ العناية لتفادى التحديات السبع المعضِلات التي ذكرناها آنفًا أمام المقارَبةِ التحليلية «لِلُّغات الطبيعية»، ولذلك تُعرَفُ وتُعرَّفُ لُغَاتُ البرمجة بأنها «لُغاتٌ مهذَّبةُ التعريف Well Defined Languages» ويمكن وِفْقَ مِنْهجيةٍ واضحةٍ صياغةُ خُوارِزْماتٍ تستطيع تحليلها تحليلاً كاملاً دون أي التباس.

٤. إلى أين وصلت المعالجَةُ الآليةُ لِلُّغات الطبيعية؟

يحاول (الجدول ٣) أدناه رَسْمَ صورةٍ تقريبيةٍ موجزةٍ لما عليه حالةُ الأبحاث والصناعة في نهاية العِقْد الثاني من الألفية الميلادية الثانية في حقل «المعالجَة الآلية لِلُّغات الطبيعية»؛ حيث نُسَمِّي عند كلِّ مستوًى من مستوياتِ اللغةِ بعضَ أهم مساحات تطبيقاتها، ثم نقوم برصد درجةِ أداءٍ [١] لأفضل الأعمال [٢] التي أُنْجِزَت في كل مساحة تطبيقية، ونذكر معه نوعَ المقارَبةِ التي اتبعها. ولربما كان أهم مَلْمَحَيْنِ يسترعيان الانتباهَ في هذه الصورة هما: أن هناك علاقة عكسية بين كُلٍّ من «درجة الأداء performance» و «الطبَقَة اللسانية (أو المستوى اللغوي) language layer» فإذا زادت إحداهُما انخفضت الأخرى، وأن مقارَبةَ «التعلم الحاسوبي» حاضرةٌ دومًا إما بمفردها أو بالتهجين مع «المقارَبة التحليلية».

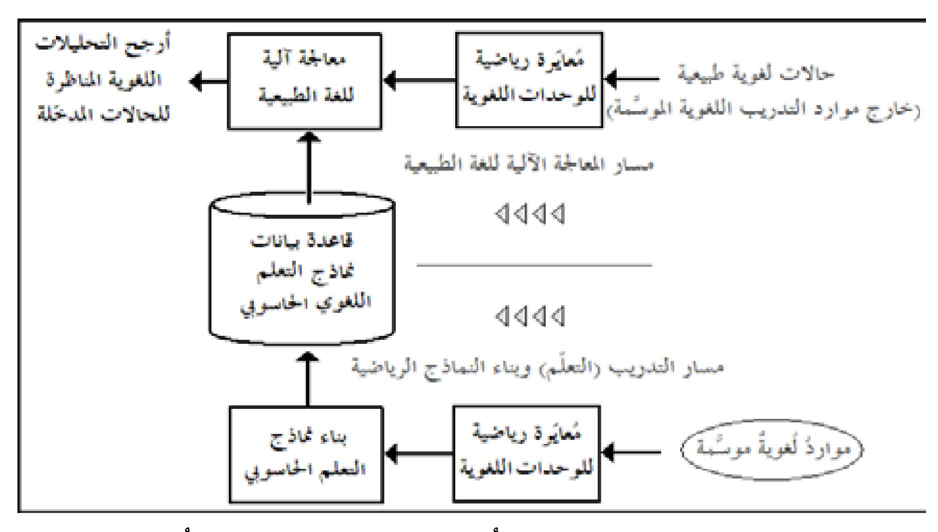
١- وذلك على مقياسٍ تنازليٍّ من خمس درجات: {صناعيٌّ جيد، صناعيٌّ مقبول، بحثيٌّ متقدِّم، بحثيٌّ أوَّليٌّ، بعيد المنال}.

٢- وذلك بصرف النظر عن اللغة (أو اللغات) الطبيعية التي يخدمها هذا العمل -وإن كانت اللغة الإِنْغِليزيةُ بحكم الأمر الواقع في وقتنا هذا هي التي تحتل مركزَ الصدارة وبؤرةَ الاهتمام.

المستوى اللغوي	المساحة التطبيقية	درجة الأداء	المقارَبة
المرسوم Orthographic	التعرف على النص المطبوع	صناعيٌّ جيد	التعلم الحاسوبي
	التعرف على النص	صناعيٌّ جيد	التعلم الحاسوبي
المنطوق Phonological\ Speech	التعرُّف على المتكلِّم	صناعيٌّ جيد	التعلم الحاسوبي
	التعرف على الكلام المنطوق	صناعيٌّ جيد	التعلم الحاسوبي
	الإملاء الآلي	صناعيٌّ مقبول	التعلم الحاسوبي
	التحقق من النطق الصحيح	صناعيٌّ جيد	التعلم الحاسوبي
	توليد الكلام المنطوق من المكتوب	صناعيٌّ جيد	هَجينة
الصرفي Morphological	التحليل الصرفي	صناعيٌّ جيد	هَجينة
	الضبط الصوتي للمفردات (التشكيل)	صناعيٌّ جيد	هَجينة
	استخلاص الصفات النحوية للمفرَدات	صناعيٌّ جيد	هَجينة
	البحث النَّصِّيّ	صناعيٌّ جيد	هَجينة
	التنقيب في النصوص	صناعيٌّ جيد	التعلم الحاسوبي
	اختصار/ تلخيص الوثائق النصية	صناعيٌّ مقبول	التعلم الحاسوبي
	مقارَنة الوثائق النصية	بحثيٌّ متقدِّم	التعلم الحاسوبي
المعجَمي Lexical	التحليل الدِّلالي المعجَمي	بحثيٌّ متقدِّم	هَجينة
	صِناعة المُعْجَمات، وإتاحتها للاستخدام	صناعيٌّ جيد	هَجينة
	أنظمة تعليمية لغوية تفاعلية	صناعيٌّ جيد	هَجينة
النحوي Syntactic	التحليل النحوي	بحثي أولي	هَجينة
الدِّلالي Semantic	التحليل الأونطولوجي	بحثي أولي	هَجينة
	الترجمة بمساعدة الحاسوب	صناعيٌّ جيد	التعلم الحاسوبي
	الترجمة الآلية	صناعيٌّ مقبول	التعلم الحاسوبي
	استخلاص توجهات الرأي العام	بحثيٌّ متقدِّم	التعلم الحاسوبي
التداولي Pragmatic	إدارة الحوار الآلي	بحثي أولي	هَجينة
	الفهم الآلي للنص	بعيد المنال	التعلم الحاسوبي
الوعي بالعالم الحقيقي	التعبير الإنشائي الآلي	بعيد المنال	هَجينة
امتلاك الذَّاتيَّة	التعبير الآلي عن المشاعر	بعيد المنال	هَجينة

الجدول ٣: صورةٌ تقريبيةٌ موجزةٌ لحالة الأبحاث والصناعة المعاصرة في حقل المعالجَة الآلية لِلُّغات الطبيعية.

ولأن «التعلُّمَ الحاسوبيَّ» حاضرٌ دَوْمًا في هذه الصورة المعاصرة للمعالجة الآلية لِلُّغات الطبيعية، فمن المفيدِ إسقاطُ إطارِهِ العامِّ المبيَّنِ في (الشكل ٧) سابقًا على «المعالجة الآلية لِلُّغات الطبيعية» وهذا ما يبينه (الشكل ١٢) أدناه؛ حيث نلاحظ أن أهم نواتج هذا الإسقاط هو أن التسمية العامة في (الشكل ٧) «العينات الغزيرة من مدخَلات النظام بالتوازي مع مخرَجاتِهِ المناظرة لها» تتحول تسميتها إلى «الموارد اللغوية الموسَّمة annotated language resources» في (الشكل ١٢)، وسوف يكون لـ «الموارد اللغوية Language Resources» و «التوسيم annotation» حضورٌ هامٌّ فيما تبقى من هذا الفصل.



الشكل ١٢: إسقاط الإطار العام لمقارَبة «التعلُّم الحاسوبي» على «المعالجة الآلية لِلُّغات الطبيعية».

قاعدة رقم ٤
يشكِّلُ «توسيم المواردِ اللغوية Language Resources Annotation» المكوِّنَ الأكبرَ في صناعةِ «المعالَجة الآلية لِلُّغات الطبيعية» المعاصرة؛ فحيث إن هذه الصناعة ترتكز أساسًا على مقارَبة «التعلُّم الحاسوبي» فيما أساليبُ وآلياتُ «التعلُّم الحاسوبي» تنظر للظاهرة التي تعمل عليها كصندوقٍ أسودَ مغلَقٍ لا يعنيها منه سوى مدخَلاته ومخرَجاته، فإن المادة المتغيرة التي ينبغي بناؤها خِصِّيصًا لكل نظام تطبيقي هي «الموارد اللغوية الموسَّمة» اللازمة لتدريبه.

٥. هل هناك سقفٌ لمعالَجة اللغات الطبيعية آليًّا؟

نعم، هناك سقفٌ للمعالَجة الآلية لِلُّغاتِ الطبيعية، وذلك لوجود سقفٍ للذكاء الاصطناعي[١] وهو الإطار العام التي تقع داخله «المعالجة الآلية لِلُّغاتِ الطبيعية» وذلك بدوره لوجود سقفٍ للرياضيات[٢] التي هي الإطار الجامع لذلك كله. فالسقف الذي يحدِّدُه سقفُ الرياضيات وسقفُ الذكاء الاصطناعي وتقع تحته آفاق «المعالجة الآلية لِلُّغات الطبيعية» يتمثل في المساحات اللغوية التي يتبدى فيها «وَعْيُ» الكاتب أو المتحدث و «ذاتِيَّتُهُ»؛ فهاتان مساحتان يتعذر مبدئيًّا على كل أدواتنا معالَجَتُهما آليًّا، ومن الحكمة عدم إهدار المجهود أو المال على البحث والتطوير فيهما أو فيما يعتمد عليهما بقوةٍ من المساحات الأخرى في هذا الحقل [٣].

وفيما تحت هذا السقف فلا يوجد من حيث المبدأ ما يقيِّد «المعالَجة الآلية لِلُّغات الطبيعية»، ولكن من حيث القدرة العملية فإن ربط وحدات اللغة الطبيعية بمختلَف مستوَياتها بالعالم الحقيقي [٤] تستلزم بناءَ «أونطولوجيا رقمية» لهذا العالم وهو ما يتطلب بدوره صياغةَ ورقمنةَ «المعرفة الفِطْرية commonsense knowledge»، وكذلك «التحليل الفِطْري commonsense reasoning»، وهي جميعًا مهامٌّ هائلةُ الضخامة [٥] لا زلنا نجهل كيف نشرع فيها ويَصْعُبُ تصوُّرُ إنجازها في العقود القريبة القادمة، وهو ما يُقِيمُ سقفًا عمليًّا للمعالجة الآلية للغات الطبيعية تحت سقفها المبدئي المشار إليه في الفقرة السابقة.

تحتَ هذين السقفين إذن تعمل «المعالَجة الآلية لِلُّغات الطبيعية»، وبما أن أنظمة تطبيقاتها العملية/ الواقعية المعاصرة لا تستغني أبدًا عن مقارَبة «التعلم الحاسوبي» - كما بَيَّنَ القسم السابق من هذا الفصل - فمن المهم أن نتعرف على كيفية أداء هذه المقارَبة

١- وقد سبقت مناقشةُ ذلك في ختام القسم الفرعي ١,٨ من هذا الفصل.

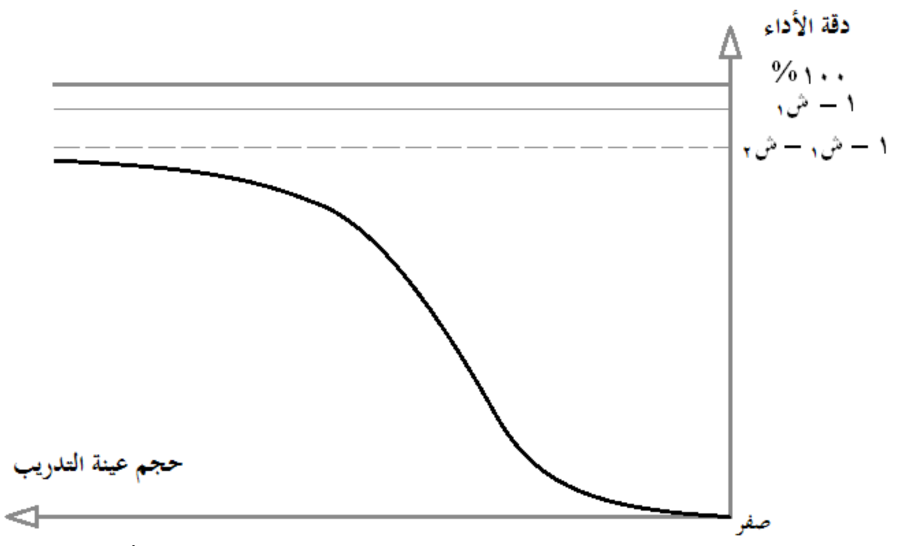
٢- وقد أشرنا لذلك في القسم الفرعي ١,٣ ثم في القسم الفرعي ١,٦ من هذا الفصل.

٣- وذلك مثل التعبير عن «المشاعر sentiments».

٤- والذي من دونه تبقى «المعالَجة الحاسوبية لِلُّغاتِ الطبيعية (NLP) Natural Language Processing» أقرب لمعالجة الرموز أو «المعالَجة الرقمية للإشارات (DSP) Digital Signal Processing».

٥- وقد سبق التعرض لهذه المعضِلات في القسم الفرعي ١,٧ من هذا الفصل في الهامش المتصل بالشكل رقم ٤، ثم أشرنا إليها مرةً أخرى مع ختام القسم الفرعي ١,٨، ثم مرة ثالثة في القسم رقم ٣ بُعَيْدَ الشكل رقم ١٠.

تَحْتَهُا. فكما بَيَّنَ (الشكل ١٢)؛ يجري تدريب «نظام المعالجة الآلية للغة الطبيعية» على «موارد لغوية موسَّمة» تمثل عيناتٍ كبيرةً متنوعةً بما يكفي من الظاهرة اللغوية المطلوب تعلُّمُها، فتتشكل نتيجةً لذلك (معامِلاتُ) «نماذج التعلم الحاسوبي»، التي تستخدمها آليةُ «المعالجة الآلية لِلُّغة الطبيعية» في «وقت التشغيل» من أجل استنباط أرجح الحلول لكل حالة لغوية تعرض عليه في إطار الظاهرة اللغوية التي جرى تعلُّمُها وذلك بدقةِ أداءٍ ق.



الشكل ١٣ : «منحنى التشبُّع» لوصف السلوك العام لعملية التدريب في مقارَبة «التعلُّم الحاسوبي»

ومع زيادة حجم «الموارد اللغوية الموسَّمة» [١] أثناء عملية التدريب، تتحسن دقة أداء المعالجَة اللغوية الآلية أثناء وقت التشغيل، لكن اطِّرادَ هذا التحسن في دقة الأداء يتباطأ بعد حجمٍ معيَّنٍ للموارد اللغوية الموسَّمة [٢] ويدخل فيما يسمى بمرحلة «التشبُّع saturation» ثم يتقارب بعد ذلك من قيمةٍ ثابتةٍ لا يتخطاها مهما زاد حجم «الموارد اللغوية» المستخدَمة للتدريب، ولا تساوي هذه القيمة الواحد الصحيح (أي لا تساوي

١– مع اشتراط المحافظة على تنوُّعها.

٢– يعتمد هذا «الحجم المعيَّن» المعروف باسم «كمية التشبُّع» على طبيعة الظاهرة اللغوية المطلوب تعلُّمها آليًّا؛ فيزيد ذلك الحجم كلما زاد تعقيد الظاهرة في ديناميكياتها وتشابكًا في علاقاتها، وعادةً ما يقع ذلك كلما علت «الطبقة اللسانية» التي تنتمي إليها الظاهرة. وجديرٌ بالذكر أن أحد أهم أسباب اتباع «مقارَبةٍ هجينةٍ» في بعض نظم «المعالجَة الآلية للغات الطبيعية» هي أن المعالجَة الجزئية عبر «المقارَبة التحليلية» قد تقلل كثيرًا «كمية التشبع» من «الموارد اللغوية الموسَّمة» المطلوبة للتدريب مما يحسن من اقتصاديات بناء النظام كُلِّه.

مئةً في المئة) بل تصغره بمقدار «هامش الخطأ البشري **ش**$_1$» مضافًا إليه «هامش خطأ التعلم الحاسوبي **ش**$_2$»، وهو ما يلخِّصُه (الشكل ١٣) أعلاه.

فأهل أية لغة من البشر يخطئون ولا بُدَّ فيها [١] ولو بهامشٍ بسيطٍ (= **ش**$_1$) ينعكس في توسيم أي مورد لغوي لتدريب أنظمة تعلُّمِها حاسوبيًّا ولا يمكن أن يتجاوز أداءُ أية آلية للتعلُّم الحاسوبي دِقَّةَ مواردها التعليمية. كما أن الغياب الاضطراري لتمثيل أركانٍ أساسيةٍ مثل «تمثيل المعرفة الفطرية» و «الارتباط بالعالم الحقيقي» و «التحليل الفطري» و «غياب السياقات» عن المادة اللغوية قيد التعلم والتحليل، و «تجاهل الطبقات اللسانية الأعلى عند معالَجة مسألةٍ لغوية من طبقة أدنى» ... إلخ ينعكس حتمًا على طَوْرَي التدريب والتشغيل كليهما ويظهر على هيئة هامش خطأٍ (= **ش**$_2$) تعتمد قيمته على أثر كل هذه الأركان الغائبة في المسألة اللغوية المعيَّنة المعنيَّة.

٦. هل لِلُّغة العربية خصوصية مع الذكاء الاصطناعي؟

تشيع في أدبيات معالَجة اللغة العربية وحوسبتها مقولاتٌ تقطع بخصوصية اللغة العربية وتفرُّدِها عن بقية لغات العالم الطبيعية الكبرى بدرجةٍ لا نظير لها من الصعوبة على جميع مستوياتها اللسانية؛ صوتيًّا وصرفيًّا ونحويًّا ودِلاليًّا ... إلخ لدرجةٍ باتت معها تلك المقولات في مقام المأثورات التي تتوارثها الأجيال وكاد تصدير الأبحاث والدراسات بها - تعليلاً للتقصير السابق في معالَجتِها و/أو تسويغًا للعمل الذي يتصدى له المؤلف - أن يكون في مقام العرف السائد الذي تُستهجَنُ مخالَفتُه. ولكن توسيع النظر في واقع اللغات الطبيعية الكبرى حول العالم يكشف الكثيرَ من التزيُّد في مثل تلك المقولات؛ فلئن كان نظامُ بناء الجملة (النَّحْو) العربي - على سبيل المثال - شديدَ المرونة يسمح بالتقديم والتأخير والحذف والتقدير ... إلخ مع ما يطرحه ذلك من صعوبات والتباسات، فإن الصينية لغةٌ غير هجائية [٢] كما أن المفردات في الصينية لا تتمايز فقط برسمها وأصواتها بل من اللازم كذلك مراعاةُ التنغيم والنبر للتمييز بينها، ونظام بنية الكلمة (الصَّرْف) في الألمانية يسمح ببناء كلماتٍ مركَّبةٍ شديدة التعقيد،

١- المقصود هنا أبناء اللغة الذين يتحدثونها بطلاقة الخبراء بها، والمقصود بأخطائهم السهو والخلط اللذان لا مَفَرَّ من وقوعهما ولو بنسبة بسيطةٍ لا تساوي صفرًا.

٢- المقصود هو لغةُ «الماندَارين» النسخةُ الرسميةُ السائدةُ من اللغة الصِّينيَّة التي لا تتكون كلماتُها من حروفٍ أبجديةٍ/ ألفبائيةٍ.

واليابانية لا تترك فواصل أو فراغات بين كلماتها، والإنجليزية لغة هجينة إلى حدٍّ بعيدٍ جدًّا على مستوى مفرداتها كما تتميز بمعدل نموٍّ معجَميٍّ مرتفع للغاية، ولغة الهَوْسا في غرب أفريقيا (مثل نيجيريا) بها أصواتٌ غير مألوفةٍ في غيرها كالأصوات الشفطية، وأغلبُ المخطوط في الفرنسية لا يُنْطَقُ وبعض ما لا يُنْطَقُ يظهر نُطْقُهُ تحت شروطٍ خاصَّة ... إلخ. والخلاصة المقصودة من ذلك هي أن الصعوبات اللغوية الظاهرة لا تقتصر على العربية، فكل لغة من لغات العالم الحية الكبرى لها صفاتها المخصوصة التي تمثل صعوباتٍ في عيون غير الناطقين بها.

وبالرغم من الاختلافات البَيِّنة بين اللغات الطبيعية على مستوى الطبقات اللسانية الدُّنْيَا والوسطى (الصوتية، والخِطاطيَّة، والصرفية، والمعجَمية، والنحوية) التي تُعْتَبَرُ اختلافاتٍ في الترميز والصياغة، فإن اللغات الطبيعية على مستوى الطبقات اللسانية العليا (الدلالي، والمجازي، والتداولي) تعكس في جوهرها وفحواها المشتركَ الإنساني من (معرفةٍ فِطْريةٍ بالعالم الحقيقي، وتحليلٍ فِطْري، ...).

ولذلك إذا اتُّبِعَت «المقارَبة التحليلية» سبيلاً في «تحليل لُغَةٍ طبيعيةٍ» فإن مَنْبَعَ الصعوبة الحقيقية هو تلك الطبقات اللسانية العليا (الغائبة غالبًا في عملية التحليل) وما يتصل بها من تعقيد الفِكْر والنفس الإنسانية [١] وليس مَنْبَعُها الطبقات اللسانية الأدنى المعنية بالترميز والصياغة؛ أما ما يبدو من صعوباتٍ ظاهريةٍ في خصوصيات كل لُغَةٍ طبيعيةٍ على تلك المستويات الترميزية والصياغية [٢] فبمقدور أهل تلك اللغة في نهاية المطاف الإحاطةُ بها والسيطرةُ عليها تستوي في ذلك العربية مع أخواتها من ألسنة البشر.

أما إذا ما اتُّبِعَت مُقارَبةُ «التعلُّم الحاسوبي» في معالجة اللغة العربية فإن تلك المقارَبة – كما أوضحنا سابقًا في نهاية القسم رقم ٤ ثم في القسم رقم ٥ من هذا الفصل – لا تفرِّق

١- راجع التحديات السبع المعضِلات أمام نمذجة الظاهرة اللغوية الطبيعية بكُلِّيَّتِها المفصَّلة في القسم رقم ٣ من هذا الفصل.

٢- في الواقع أن العربية إن كانت لها خصوصية في هذا الصدد بين ألسنة البشر الحية فهي أنها أكثرُها نظامًا؛ خذ مثالاً الاتساقَ الكبير بين نطق النص ورسمه، ومثالاً آخَرَ غلبةَ انتظام الاشتقاق الصرفي العربي من عدد صغير نسبيًّا من المواد وعددٍ أصغر من الصِّيَغ الصرفيةِ، ومثالاً ثالثًا اتساقَها الدِّلاَلي المعجَمي إلى حدٍّ بعيد، ومن ثم قارِنْهُ مع ضعف أو غياب مثل ذلك الانتظام في كثيرٍ من مظاهر اللغات الأخرى – راجع المعجَم الإنغليزي المعاصر على سبيل المثال – وهي أطولُها اتصالاً في التاريخ (عبر حوالي أَلْفَيْ عام) وهي أكثرُها توثيقًا وتقعيدًا عبر تاريخها الطويل المتصل. وإن كان هناك من أثرٍ لتلك الخصوصيات في مظاهر اللغة العربية فلا بُدَّ أن يكون أثرًا إيجابيًّا للغاية عند اتباع أية مقارَبةٍ تحليليةٍ لها على نقيض ما تزعم المقولات الشائعة التي بدأنا القولَ بها.

بين اللغة العربية وأية لغةٍ طبيعيةٍ أخرى؛ حيث لا تعنيها التفاصيلُ الداخليةُ للظواهر (اللغوية) التي تقوم بتعلُّمِها فهي تنظرُ إليها جميعًا كصندوقٍ أسودَ مغلَقٍ لا يميِّزُها عن بعضها البعض سوى الموارد اللغوية الموسَّمة اللازمة لتدريب آلية «التعلُّم الحاسوبي». [1]

قاعدة رقم ٥
إن تعليل التقصير والتأخر في حقل حوسبة اللغة العربية وتطبيقات المعالجَة الآلية لها بدعوى اتصاف اللغة العربية لخصوصيَّاتٍ تجعل من حوسبتها ومعالجَتِها الآلية أمرًا بالغ الصعوبة مقارنةً مع غيرها من اللغات الطبيعية لَهُوَ مَحْضُ وَهْمٍ ينبغي تَنْحِيتُه؛ فدرجة الصعوبة في حوسبة اللغة العربية ومعالجَتِها آليًّا لا تختلف عن صعوبة حوسبة أية لغة طبيعية أخرى ومعالجتها آليًّا، سواء اتبعت في تلك المعالجَة أساليبُ تحليليةٌ أو أساليبُ للتعلُّم الحاسوبي أو أساليبُ هجينةٌ بينهما.

٧. ما هي آفاق مستقبل المعالجة الآلية لِلُّغة العربية عبر الذكاء الاصطناعي؟

مثلها مثل سواها من اللغات الطبيعية الكبرى في العالم، فإن آفاق «المعالجة الآلية للغة العربية» تقع تحت السقفين اللذين جرى تعريفُهُما في القسم رقم ٥، ويمكن لها خلال ما لا يجاوز عقدًا من الآن (أي بنهايات العِقْد الثالث من الألفية الميلادية الثانية) أن تتدارك ما فاتها كي تلحق بالصورة – التي رصدها سابقًا الجدولُ رقم ٣ في القسم الرابع من هذا الفصل – لحالة الأبحاث والصناعة المعاصرة في حقل المعالجَة الآلية لِلُّغات الطبيعية، فتتحقق بذلك مكاسبُ اقتصاديةٌ كبيرةٌ جَرَّاءَ المنافسة في هذا القطاع العظيم من اقتصاد المعلومات والمعرفة، وتتحقق آثارٌ ثقافيةٌ إيجابيةٌ عميقةٌ سوف تمثل دفعةً قويةً للقوة

١- ولذلك تحرص كبرَيات شركات التقنية المعلوماتية في عصرنا الراهن التي تنتج نظمًا لمعالجة اللغات الطبيعية تغطي مختلَف (أو عديدًا من) اللغات الطبيعية الحية على اتباعَ مقارَبةَ «تعلُّم حاسوبيٍّ» صِرْفةٍ وتحاول تفادي أية «مقاربات تحليلية» ولو كانت جزئيةً؛ حيث يكلِّفها ذلك إنتاج نظام معالجَةٍ واحدٍ مع إنتاج موارد لغوية متعددة بتعدد اللغات المختلفة التي يغطيها، وذلك بدلاً من إنتاج أنظمة معالجَةٍ متعددةٍ بتعدد تلك اللغات المختلفة. خذ مثالاً على ذلك «خدمة غووغِل الشهيرة للترجمة» التي تغطي عددًا = ن من اللغات الطبيعية، فمع مقارَبة التعلُّم الحاسوبي – وهي المتَّبَعة فعلاً – فإن نظام الترجمة واحد وما يتعدد فهو الموارد اللغوية اللازمة لتدريبه على الترجمة بين كل زوج من أزواج اللغات المشمولة البالغ عددها ن×(ن-١) من هذه الأزواج، فإذا فكرت «شركة غُووغِل Google» أن تتبع مقارَبةً تحليليةً لاحتاجت إلى تطوير ن×(ن-١) من الأنظمة المختلفة كي تغطي الترجمة بين كل أزواج اللغات المشمولة.

الناعمة العريقة لهذه الأمة، وتتحقق أيضًا نجاحاتٌ علميةٌ وتقنيةٌ تسهِم في النهوض بالبحث والتطوير في بلادنا في حقل الحوسبة والمعلوماتية على وجه العموم.[1]

قاعدة رقم ٦
من أجل أن تَلْحقَ «المعالجَةُ الآلية لِلُّغةِ العربية» خلال نحو عِقْدٍ من الزمان بأفضل ما وصلت إليه أبحاث وصناعة «المعالجة الآلية لِلُّغات الطبيعية»، يلزمُ إجراءُ ما يلي:
أولاً: إعداد جيلٍ من الأساتذةِ النابهين في «هندسة وعلوم الحاسب» الخبراءِ في الذكاء الاصطناعي بمدرسته المعاصِرة لجعلهم على إلمامٍ ودرايةٍ جيدةٍ بأسس علوم اللغة واللسانيات العربية وتحدياتها.
ثانيًا: إعداد جيلٍ من الأساتذةِ النابهين في «علوم اللغة واللسانيات العربية» لجعلهم على إلمام ودراية جيدة بمبادئ «علوم الحاسب» ونُظُمِه، ومبادئ «الذكاء الاصطناعي» وقضايا حوسبة اللغة الطبيعية ومعالجتها آليًّا.
ثالثًا: أن يؤهلَ الصنف الأول من الأساتذة قاعدةً أوسع من طلبة هندسة وعلوم الحاسب النابهين بنفس الكيفية التي جرى إعداد هؤلاء الأساتذة بها.
رابعًا: أن يؤهلَ الصنف الثاني من الأساتذة قاعدةً أوسع من طلبة علوم اللغة واللسانيات العربية النابهين بنفس الكيفية التي جرى إعداد هؤلاء الأساتذة بها.
خامسًا: تقوم على هذه القاعدة الحاسوبية اللسانية التي جرى بناؤها في الخطوات السابقة صناعةُ «الموارد اللغوية العربية وتوسيمها» كصناعةٍ معلوماتيةٍ ثقيلةٍ، كثيفةِ العمالة.
سادسًا: لا بُدَّ أن يَضُمَّ كُلُّ مشروع لصناعة «الموارد العربية اللغوية وتوسيمها»، و/أو توظيفها في مقارَبة «تعلُّم حاسوبيّ» أو «مقارَبةٍ هجينةٍ» لصناعة تطبيقٍ على «المعالجَة الآلية لِلُّغة العربية» الأصنافَ الثلاثة الآتية من العاملين بالأوزان المناسبة لحاجة المشروع: المتخصصين في علوم اللغة واللسانيات العربية، والمتخصصين في هندسة وعلوم الحاسب ولا سيما الذكاء الاصطناعي، و«محترفي إدارة المشروعات Project Management Professionals (PMPs)»[1].

ولعل هذا الفصل وهذا الكتاب أن يكونا خطوةً على هذا الطريق.

---

١ – ويقوم هؤلاء بتطبيق آليات التخطيط والتنفيذ والمتابعة التي تساعد هذه المشروعات «كثيفة العمالة» و«متشابكة المهام» و«ذات الحساسية العالية للدقة» على الوصول لبر الأمان؛ أي وفق مواصفات الجودة، ودون تجاوز الميزانية ولا الإطار الزمني.

٨. الخاتمة والنتائج

ارتحلنا عبر هذا الفصل في تاريخ حلم محاكاة الآلات للذكاء البشري منذ التاريخ القديم بالتدرُّج مع تطوُّر الأدوات المُمَكِّنةِ له خلال القرون القليلة الماضية، ثم فَصَّلنا في ظروف تأسُّس حقل «الذكاء الاصطناعي» في العصر الحديث ولا سيما مع بدء بناء الحواسيب الرقمية واستعرضنا بالتفصيل المراحل التي مر بها صعودًا وهبوطًا إلى أن استقر على مسار تصاعُدِيٍّ سريعٍ منذ حوالي ربع القرن.

قدم هذا الفصل بعد ذلك المقارَبَتَيْن الأساسيتين لحل المسائل المعتَبَرة في إطار الذكاء الاصطناعي، والتحدياتِ الأساسيةَ أمام قضية «المعالَجَة الآلية لِلُّغات الطبيعية» وهي إحدى المباحث الكبرى للذكاء الاصطناعي، ومضى الفصل ليبيِّن حالَ هذه القضية مع كل مقاربةٍ منهما، ولماذا مالت الكفة نحو مقارَبة التعلُّم الحاسوبي» مقابلَ «المقارَبة التحليلية»، ثم انتقلنا بعد ذلك لرسم صورة الحالة الواقعية الراهنة لتطور الأبحاث والصناعة في حقل «المعالَجَة الآلية لِلُّغات الطبيعية»، وبعدها أجبنا على سؤال السقف الأعلى لما يمكن أن تبلغه الأبحاث والصناعة في هذا الحقل – على الأقل في المستقبَل المنظور.

وأخيرًا انتقل التركيز على إسقاط ما سبق على «المعالَجَة الآلية لِلُّغة العربية» وهل هناك حقًّا خصوصياتٌ لها تجعلها أعصى على المعالَجَة الآلية من اللغات الأخرى؟ وما هي الآفاق المستقبلية لحوسبة العربية ومعالجتِها آليًّا؟

وفي الختامِ نؤكِّد أن هذا الفصل – رغم كل ما يحتشد به تفاصيلَ تاريخيةٍ وعلميةٍ – ليس في مقام تقديم شرحٍ أو تحليلٍ تِقْنِيٍّ مستوْفٍ لأية آليةٍ رياضيةٍ أو حاسوبيةٍ من تلك التي يعرضها، فتلك الشروح والتحليلات وفيرةٌ مستقرةٌ في بطونِ الكثير والكثير من الأسفار المتخصصة الموجَّهةِ إلى المتخصصين في علوم الحاسب والرياضيات المتقدمة. ولكن رسالةَ هذا الفصل تتوجه أساسًا إلى العاملين النابهين في علوم اللغة العربية وما يَرْفِدُها من حقولٍ [1] ممن يتطلعون و/ أو يتصدون بالفعل لتوظيف الحوسبة وتقنياتها

١– وتتوجه أيضًا هذه الرسالة بطبيعة الحال إلى كل مثقَّفٍ عربيٍّ مهتَمٍّ بالإلمام بماهيةِ الذكاءِ الاصطناعيِّ وتاريخِهِ وواقع ما يوفره من آلياتٍ وتقنيَّات لمعالجةِ اللغاتِ الطبيعيةِ؛ وخصوصًا العربية، وحدودِ إمكانياتِ هذه الآليات والتقنيات وآفاقِها.

في تطوير معالجة اللغة العربية لبناء تطبيقات ومشروعات ذات مردودٍ تربويٍّ و/ أو ثقافيٍّ و/ أو معلوماتيٍّ و/ أو اقتصاديٍّ، وذلك بُغْيَةَ تعريفهم بوضوحٍ بجذور وواقع هذا المجال في إيجازٍ وتبسيطٍ غير مُخِلَّيْنِ مما يساعدهم على وضع أهدافٍ واقعيةٍ لتطبيقاتهم ومشروعاتهم واتخاذ خياراتٍ صحيحةٍ تؤدي إلى إنجاز هذه الأهداف بكفاءةٍ وجودةٍ عاليتَيْن.

ببليوجرافيا مرجعيَّة.

1. Aggarwal, C.C. Springer (2018) *Machine Learning for Text, Springer*.
2. Buchanan, B.G. (2006: Winter 2005) *A (Very) Brief History of Artificial Intelligence*, AI Magazine, AI Magazine: Volume 26 – Number 4, pp. 53–60, USA.
3. Dominigos, P. (2015) *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*, Basic Books, New York - USA.
4. Goodfellow, I., Bengio, Y., Courville, A., (2016) *Deep Learning*, MIT Press, Massachusetts – USA.
5. Hofstadter, D. (1979) *Gödel, Escher, Bach: an Eternal Golden Braid*, Basic Books, New York USA.
6. Indurkhya, I. (Editor), Damerau, F.J. (Editor), (2010) *Handbook of Natural Language Processing* (2nd edition), Chapman & Hall Book/ CRC Press, Florida – USA.
7. Jurafsky, D., Martin, J. H., (2008) *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Processing*, (2nd edition), Prentice Hall, USA.
8. Lane, H., Howard, C., Hannes, M.H., (2019) *Natural Language Processing in Action,* Manning Publication, USA.
9. Mitkov, R. (editor), (2005) *The Oxford Handbook of Computational Linguistics*, Oxford University Press – UK.
10. Nilsson, N.J. (2015) *The Quest for Artificial Intelligence*, Cambridge University Press, UK.
11. Russell, S., Peter Norvig, P., (2009) *Artificial Intelligence: A Modern Approach*, (3rd edition), Prentice Hall, USA.
12. Schutze, H., Manning, C.D., (2000) *Foundations of Statistical Natural Language Processing*, MIT Press, USA.

الفصل الثَّاني

المُعالَجة الآليَّة للُّغة العربيَّة المكتوبة

مُقدِّمة في ذكاء الآلة

د. الـمُعتزّ بالله السَّعيد

ملخص

يُقدِّمُ هذا الفصلُ مدخلًا إلى مُعالجَة اللُّغة العربيَّة المكتوبة عبرَ بوَّابة الذَّكاء الاصطناعيّ، في مُحاولةٍ للإجابة عن السُّؤال البحثيّ: كيفَ يُمكنُ للآلة أن تُحاكي ذكاءَ الإنسان في فهم اللُّغة العربيَّة ومُعالجَتِها في بنيتِها المكتوبة؟. ويعرضُ الباحثُ لذلكَ في مُقدِّمةٍ وأربعة محاورَ أساسيَّةٍ تُمثِّلُ مُستويات مُعالجَة اللُّغة المكتوبة، وهي (مُستوى الـمُعالجَة الكتابيَّة، ومُستوى الـمُعالجَة الصَّرفيَّة، ومُستوى الـمُعالجَة التَّركيبيَّة، ومُستوى الـمُعالجَة الدِّلاليَّة). ويعرضُ الباحثُ في كُلِّ مُستوًى لوحدة الـمُعالجَة (التَّحليل)، والتَّوجيه الآليّ للمُعالجَة السَّطحيَّة والعميقة لها، وأبرز تطبيقاتها في اللُّغة العربيَّة.

ويستهدفُ الفصلُ القارئَ العربيَّ ذا الخلفيَّة اللُّغويَّة الَّذي ينشُدُ التَّعرُّفَ على طرائق الـمُعالجَة الآليَّة للُّغة العربيَّة المكتوبة ومناهجِها، على النَّحو الَّذي يفتحُ أمامَهُ الطَّريقَ إلى فهمٍ أكثرَ عُمقًا. لهذا، يسعى الباحثُ إلى عَرضِ أفكاره انطلاقًا من منطق اللُّغة العربيَّة [القائم على الكَيف، عَبرَ السَّماع والقياس]، مع مُوائمةٍ مُبَسَّطةٍ بمنطق الآلة [القائم على الكمّ، عَبرَ الـمُحاكاة الرِّياضيَّة]. ويلتزمُ الباحثُ منهجًا وصفيًّا تحليليًّا، يسعى من خلالِهِ إلى وصف واقع اللُّغة العربيَّة المكتوبة، وتحليلها في مُستوياتِها الـمُختلفة، بُغيَةَ الوُقُوف على أساليب توجيه الآلة إلى فهمِها ومُعالجَتِها على الوَجه الـمُوافق لطبيعتِها.

الكلمات المفتاحية:

مُعالجَة اللُّغـــات الطَّبيعيَّة Natural Language Processing (NLP)، اللُّغـة العربيَّة المكتوبة Written Arabic ، الذَّكاء الاصطناعيّ (Artificial Intelligence AI)، تقنيات اللُّغات الإنسانيَّة (Human Language Technology (HLT.

١ . الذَّكاء الاصطناعيّ واللُّغة المكتوبة.

اللُّغةُ المنطوقةُ أسبقُ إلى الوُجود من اللُّغة المكتوبة؛ إذ لم يبتكر الإنسانُ أبجديَّاتِ الكتابة إلَّا في حِقبةٍ زمنيَّةٍ مُتأخِّرةٍ نسبيًّا عن حِقبةِ وُجُوده. ومع ظُهور الحواسيب وتطوُّرها في مُنتَصَف القرن العشرين، ظهرَت مُحاولاتٌ وتجاربُ أوَّليَّةٌ تهدفُ إلى إخضاع الآلة لـمُحاكاة ذكاء الإنسان في فهم اللُّغات الطَّبيعيَّة وتحليلها؛ إلَّا أنَّ هذه الـمُحاولاتِ والتَّجاربَ قد انطلَقَت من اللُّغة المكتوبة، لأسبابٍ عديدة، أهمُّها: الطَّبيعة القياسيَّة النِّسبيَّة لها، وإمكانيَّة التَّفاعُل مع بنيتِها الملموسة، وسُهولة حفظِها واسترجاعِها، وثبُوت صُورتِها الخارجيَّة بصرف النَّظَر عن مُستخدميها أو الـمُتفاعلينَ معها.

يتناولُ هذا الفصلُ مُحاولةً للإجابة عن السُّؤال المحوريّ بشأن تعاطي الآلة مع اللُّغة الطَّبيعيَّة في صُورتِها المكتوبة؛ أي: كيفَ يُمكنُ للآلة أن تُحاكي ذكاءَ الإنسان في فهم اللُّغة الطَّبيعيَّة المكتوبة؟ ولا ننشُدُ الحديثَ عن عُمقِ الفكرة وإجراءاتِها الَّتي قد تبدو مُعقَّدةً إلى حدٍّ ما؛ لا سيَّما في أبعادِها الرِّياضيَّة والهندسيَّة الـمُتشعِّبة؛ لكنَّنا نلتمسُ تقديمَ طرحٍ تمهيديّ، يُمكنُ من خلاله بناءُ قاعدةٍ مُستقبليَّةٍ، تكونُ مُنطلَقًا لتناوُلٍ أكثرَ عُمقًا لـمُعالَجة اللُّغات الطَّبيعيَّة (Natural Language Processing (NLP. وسيكونُ حديثُنا عن اللُّغة الطَّبيعيَّة المكتوبة بصُورةٍ عامَّة، مع التَّركيز على اللُّغة العربيَّة.

حينَ نسعى إلى إخضاعَ الآلة لفهم اللُّغة الطَّبيعيَّة، فعلينا أن نُقَدِّرَ طبيعةَ الـمُعطَيات اللُّغويَّة الـمُوَجَّهة للآلة ومدى قابليَّة هذه الـمُعطَيات للتَّحليل الكمِّيّ. ذلكَ أنَّنا نخلقُ بيئةً تُحاكي قُدرةَ الإنسان على استخدام اللُّغة الطَّبيعيَّة؛ وهذه القُدرةُ لم تَصنَعها قُوى الطَّبيعة؛ لكنَّها جاءَت نتيجةَ عمليَّاتٍ مُتدرِّجة لاكتساب المهارات اللُّغويَّة (الاستماع، والتَّحدُّث، والكتابة، والقراءة). فالطِّفلُ النَّاشئُ في مُجتَمَع يستخدمُ لُغةً مُعَيَّنةٍ يستمعُ إلى أصوات اللُّغة في مُحيط مُجتمَعِه، وينطقُ في مراحل النَّشأة الأولى بعضًا من هذه الأصوات، ثُمَّ يُؤلِّفُ بينَها في كلماتٍ، ويصنعُ جُملًا بسيطةً ومُركَّبة، ثُمَّ يتعلَّمُ كتابةَ المحارف وقراءَتَها، ثُمَّ الكلمات والجُمَل، وهكذا، حتَّى يصيرَ قادرًا على استخدام اللُّغة في بيئةٍ تفاعُليَّة. وترتبطُ هذه القُدرةُ بالـمُعطَيات اللُّغويَّة الَّتي توافرَت له؛ فإذا نشأَ مثلًا في مُجتمَع ذي ثقافةٍ لُغويَّةٍ واسعةٍ، فمن البَدَهيّ أن يتأثَّرَ بهذه الثَّقافة في تكوينِه، والحالُ كذلكَ إذا نشأَ

في مُجتمَع محدودِ الثَّقافة. والآلةُ قريبةٌ من ذلكَ؛ حيثُ تنشأُ علاقةٌ طرديَّةٌ بينَ قدرتِها على فهم اللُّغات الطَّبيعيَّة من ناحية، واتِّساع مدى المعارف اللُّغويَّة الـمُوَجَّهة لها من ناحيةٍ أخرى. وهذا يُفَسِّرُ وُجُودَ تبايُنٍ بينَ مُخرَجات أدوات مُعالَجَة اللُّغات الطَّبيعيَّة، سواءٌ في اللُّغة الواحدة، أم بينَ اللُّغات الـمُتعدِّدة.

يكمُنُ الإشكالُ الحقيقيُّ في وُجُود فجوةٍ بينَ الذَّكاء الطَّبيعيّ [ذكاء الإنسان] والذَّكاء الاصطناعيّ [ذكاء الآلة]. ومَرَدُّ ذلكَ إلى أنَّ مُحاكاةَ الطَّبيعة لا يُمكنُ أن تحدُثَ بصُورةٍ كُلِّيَّة؛ حيثُ تظلُّ بعضُ سِمات الشَّيء الطَّبيعيّ [الأصل] خاصَّةً به، وغيرَ قابلةٍ للاستحداث في الشَّيء المصنُوع [التَّقليد]. والأمرُ يتَّصلُ بما يُمكنُ إخضاعُهُ للقياس في الشَّيء الطَّبيعيّ. وفي النَّمُوذَج الَّذي نحنُ بصَدَدِه، بشأن مُحاكاة ذكاء الإنسان في الآلة، يُمكنُ مثلًا ان نُحاكي القُدرةَ على مُمارسةٍ نشاطٍ خارجيّ، كالقُدرة على القراءة أو الكتابة؛ لكنَّنا لا نستطيعُ مُحاكاةَ المشاعر الدَّاخليَّة، كالحُبِّ والألَمِ والغَيرة. وأقصى ما نستطيعُ أن نصلَ إليه في هذا الشَّأن أن نُخضِعَ الآلةَ إلى اكتشاف وُجُود بعض المشاعر، دونَ أن تتولَّدَ هذه المشاعرُ فيها. ولا يتمُّ ذلكَ إلَّا عبرَ قرائنَ [لُغويَّة وغَير لُغويَّة] قابلةٍ للقياس.

والواقعُ أنَّ إخضاعَ الآلة لفهم اللُّغة الطَّبيعيَّة يتجاوزُ حُدُودَ اللُّغة ذاتِها إلى عواملَ خارجيَّة تتَّصلُ بطبيعة الإنسان النَّفسيَّة ومُحيطِه الاجتماعيّ. ويُوَضِّحُ (الجدول ١) مثالًا على ذلك.

م	الجُملة	السِّياق اللُّغَويّ	السِّياق الاجتماعيّ
١	انفعلَ الأبُ على ابنِه، لأنَّهُ مُشاغِب	ثابت	مُتغيِّر
٢	انفعلَ الأبُ على ابنِه، لأنَّهُ مُغاضِب		

الجدول ١ : نموذَج توضيحيّ لاختلاف السِّياق اللُّغويّ عن السِّياق الاجتماعيّ.

يحتوي الجدولُ على جُملَتَين مُتَّفِقَتَين تمامًا في السِّياق اللُّغويّ؛ وهو السِّياقُ الَّذي يُمكنُ التَّحكُّمُ فيه؛ لكنَّهما تختلفان في السِّياق الاجتماعيّ؛ وهو السِّياقُ الَّذي يَفهمُهُ الإنسانُ ويتفاعلُ معهُ بصورةٍ ديناميكيَّة. ويستدعي هذا الفهمُ إدراكَ العلاقةِ بينَ الأب والابن، لتحديد مَن يتَّخذُ سمةَ (الـمُشاغَبة) ومَن يتَّخذُ سِمةَ (الـمُغاضَبة). وسنصلُ بذلكَ إلى أنَّ:

- الجُملة ١ = الابن مُشاغِب.
- الجُملة ٢ = الأب مُغاضِب.

ومثلُ ذلكَ يُفسِّرُ الحاجةَ الدَّائمةَ إلى التَّدخُّل البشريّ في مُراجَعة مُخرَجات الآلة، مهما بلَغَ مُستوى ذكائِها الاصطناعيّ. وفي إطار سَعينا إلى استكشاف تدرُّج الآلة من مُحاكاة اللُّغة الطَّبيعيَّة المكتوبة في أبسطِ أشكالِها إلى مُحاكاتِها في أعمقِ مداها، لا سيَّما في اللُّغة العربيَّة، فسنعرضُ في الصَّفحات التَّالية أربعةَ محاورَ أساسيَّة، تُمثِّلُ مُستويات مُعالجَة اللُّغة العربيَّة المكتوبة، بدءًا بمُستوى مُعالجة محارف العربيَّة، ومُرُورًا بمُستوى مُعالجة البنية ومُستوى مُعالجَة التَّركيب، وانتهاءً بمُستوى مُعالجة المعاني.

٢. المُعالجَة الآليَّة الكتابيَّة [المحرفيَّة / الجرافيميَّة].

٢,١. وَحدة المُعالجَة الكتابيَّة.

وَحدَةُ الـمُعالجَة الكتابيَّة هي (الجرافيم Grapheme) [أو المحرَف]، الَّذي يُعَدُّ أصغرَ وَحدةٍ خطِّيَّةٍ للُّغة المكتوبة؛ ويُقابلُهُ في اللُّغة المنطوقة (الفونيم Phoneme) الَّذي يُعَبِّرُ عن الصَّوت اللُّغَويّ. وتشملُ الوحداتُ الكتابيَّة [الجرافيمات] - عُمُومًا - جميعَ محارف اللُّغات الإنسانيَّة، سواءٌ أكانَت حُرُوفًا أم أرقامًا أم رُمُوزًا رياضيَّة، أم غَير ذلك. ويُمكنُ التَّمثيلُ على الجرافيمات بحُرُوف العربيَّة (ء، ب، ت، ث)، وحُرُوف اللُّغــات اللَّاتينيَّـة [مثل: الفرنسيَّة] والجـــرمانيَّة [مثل: الإنجليزيَّة] (A, B, C, D)، والحُرُوف الصِّينيَّة (手، 安، 児، 傘)، وعلامات التَّرقيم (؟، !،: ، /)، والأرقام العربيَّة (1, 2, 3, 4)، والأرقام الهنديَّة (١، ٢، ٣، ٤)، والرُّموز الرِّياضيَّة (+، -، *، =)، وكذلكَ علامات الضَّبط الَّتي تُعبِّرُ عن الشَّكل المكتوب للأصوات المنطوقة من الحركات والسَّكَنات [في بعض اللُّغات، كالعربيَّة والعبريَّة]؛ على النَّحو الَّذي نجدُهُ في علامات الضَّبط العربيَّة (ـــُـــ، ـــَـــ، ـــِـــ، ـــْـــ).



الشَّكل ١: صُورةُ الجرافيم في اللُّغة العربيَّة

وقد يأخُذُ الجرافيمُ شكلًا كتابيًّا واحدًا أو أكثرَ من شكل، بحسب النِّظام الكتابيّ للُّغة الطَّبيعيَّة، وبحب موضِعِه في الكلمة [أو السِّياق التَّعبيريّ]. ويُعرَفُ الشَّكلُ الواحدُ للجرافيم بـ (ألوجراف Allograph). وعلى سبيل المثال، يأخُذُ جرافيم (A) في الإنجليزيَّة ألوجرافَين فقط، هما (A, a)، بينما يأخُذُ جرافيم (ء) [الهمزة] في العربيَّة خمسةَ ألوجرافات، هي (ء، أ، إ، ؤ، ئ).



الشَّكل ٢: صُورةُ الألوجراف في اللُّغة العربيَّة

٢, ٢. التَّوجيه الآليّ لمُعالجة الوحدات الكتابيَّة في اللُّغة العربيَّة.

كُلُّ حَرفٍ مَحرَف، وليسَ كُلُّ مَحرَفٍ حرفًا؛ لأنَّ المحارفَ تشملُ الحُرُوفَ والرُّموزَ الكتابيَّة الأخرى في اللُّغات الطَّبيعيَّة. ويتمُّ توجيهُ الآلة لـمُعالجَة المحارف باعتبارِها وحدات الكتابة على إحدى صُورَتَين:

١) صُورة مُفرَدة: ونعني صُورة الحُرُوف والرُّمُوز.

٢) صُورة مُركَّبة: ونعني صُورة الكلمات أو المجاميع الكتابيَّة.

وسواءٌ أجاءَت هذه المحارفُ [الوحدات الكتابيَّة] مُفرَدةً أم مُرَكَّبة، يُمكنُ إخضاعُها للـمُعالجَة الآليَّة عبرَ ما يُعرَفُ بـ (خوارزمات التَّعلُّم الآليّ Machine Learning Algorithms) الَّتي تُمثِّلُ أساليبَ منهجيَّةً لتوجيه الآلة إلى فهم طبيعة اللُّغة الطَّبيعيَّة

على أساسٍ رياضيٍّ. ويعني هذا أنَّ علينا أن نُقدِّمَ للآلة مُعطياتٍ ثابتة (مُحصاة) لتستندَ إليها في مُحاكاة ذكاء الإنسان في فهم اللُّغة. ومع وجود عددٍ كبيرٍ من مناهج التَّعلُّم الآليّ، فإنَّنا نُرَكِّزُ على منهجين رئيسَين في مُعالَجة اللُّغة العربيَّة، هما:

١) التَّعلُّم على أساس القواعد Rule-based Machine Learning.

٢) التَّعلُّم على أساس الـمُدوَّنات [الذَّخائر اللُّغَويَّة] Corpus-based Machine Learning.

وعبرَ هذين المنهجَين، يُمكنُ تقسيمُ مُستوى الـمُعالَجة ذاتِهِ إلى قِسمَين، على النَّحو الآتي:

١, ٢, ٢. المُعالَجة السَّطحيَّة للوحدات الكتابيَّة.

إذا تناولنا الوحدات الكتابيَّة العربيَّة في صُورتِها الـمُفرَدة، فسنجدُها مجموعةً من المحارف المعلومة والـمُحصاة، وإن تفاوتَ إحصاؤُها وفقًا لأسلوب الحَصر ذاتِه. وإذا تناوَلناها في صُورتِها الـمُرَكَّبة، فسنجدُها كلماتٍ معلومةً في أذهان أبناء اللُّغة وفي المعاجم العربيَّة المكتوبة. وأيًّا كانَ الأمرُ، فالـمُعالَجَةُ السَّطحيَّةُ للوحدات العربيَّة المكتوبة تقومُ أساسًا على مُطابقة المادَّة الخاضعة للـمُعالَجة بواقع اللُّغة.

وحتَّى نستبينَ ذلكَ نقول:

إذا اعتمَدنا أسلوبَ التَّعلُّم الآليّ على أساس القواعد، فالـمُعالَجَةُ السَّطحيَّةُ تستدعي مُطابقةَ الوحدات الكتابيَّة الخاضعة للـمُعالَجة بقواعد الكتابة العربيَّة الَّتي تُوَجَّهُ الآلةُ إليها عبرَ قواعد بياناتٍ حاويةٍ لها. ومن هذه القواعد مثلًا: أشكالُ حُرُوف العربيَّة في أوَّل الكلمة وأوسطِها وآخرها، وحُرُوف الزِّيادة العربيَّة، ومواطن هذه الحُروف حينَ تأتي في صُورة السَّوابق، ومواطنها حينَ تأتي في صُورة اللَّواحق، وغَير ذلك. وتبدو هذه القواعدُ بسيطةً على حدٍّ بعيدٍ؛ إذ تكونُ معلومَةً لأبناء اللُّغة، ولا يستدعي استكشافُها مُعالَجاتٍ سابقةً في أكثر الأحيان. ويُبَيِّنُ (الجدول ٢) بعضَ النَّماذج لهذه القواعد.

م	المحارف العربيَّة	القاعدة
١	ء، ؤ، د، ذ، ر، ز، و	لا تتشابكُ مع حُرُوفٍ لاحِقة
٢	ا	تأتي مُستقلَّةً في أوَّل الكلمة دائمًا
		تأتي مُستقلَّةً في وسَط الكلمة أو آخِرِها، إذا سُبِقَت بالمحارف: (ء، ؤ، د، ذ، ر، ز، و)
		تأتي غَيرَ مُستقلَّةٍ في وسَط الكلمة أو آخِرِها، إذا سُبِقت بالحُرُوف الأخرى
٣	ـة	تأتي في نهاية الكلمة، دونَ أوَّلِها أو أوسَطِها
٤	الـ	تأتي سابقةً لإفادة التَّعريف
		تأتي مُستقلَّةً أو مسبوقةً بـحُرُوفٍ مُعَيَّنة (مثل: و، ف، ك)
٥	ـها	تأتي لاحقةً للدِّلالة على ضمير الـغائبة
٦	الضَّمائر الـمُنفصلة (أنا، نحن، أنتم، ...)	تقبلُ السَّوابق
		لا تقبلُ اللَّواحق

الجدول ٢: من قواعد الكتابة العربيَّة الـمُستخدَمة في الـمُعالَجة السَّطحيَّة للوحدات الكتابيَّة

أمَّا إذا اعتمَدنا أسلوبَ التَّعلُّم الآليّ على أساس الـمُدوَّنات، فالـمُعالَجةُ السَّطحيَّةُ تستدعي مُطابقةَ الوحدات الكتابيَّة الخاضعة للـمُعالَجة بمُقابِلاتٍ لها. وتبدو هذه الـمُقابلات في صُورة كلماتٍ مُستمدَّةٍ من مُدوَّناتٍ لُغويَّةٍ Linguistic Corpora مُمثِّلةٍ لواقع اللُّغة؛ حيثُ تُعادُ هَيكلةُ نُصوص الـمُدوَّنات لتظهرَ أشبهَ بمُعجم لـمُفرَدات اللُّغة. وحالَ الـمُطابَقة، تُوَجَّهُ الآلة إلى التَّعرُّف على الكلمات الَّتي تكونُ لها مُقابلاتٌ في هذا الـمُعجم، وتُوَجَّهُ كذلكَ إلى تجاهُل ما عداها.

٢,٢,٢. المُعالَجَة العميـــــقـــــة للوحدات الكتابيَّة.

السِّمةُ الأساسيَّةُ للـمُعالَجة العميقة للوحدات الكتابيَّة في العربيَّة أنَّها تعتمدُ على مُعطَياتٍ غَير معلومةٍ مُسبقًا؛ بمعنى أنَّ مُوجِّهات الآلة تتجاوزُ الـمُعطَيات السَّطحيَّة لأشكال المحارف وسماتها الظَّاهرة إلى مُعطَياتٍ أخرى ترتكزُ على ظواهرَ لُغويَّة غَير ظاهرةٍ للمحارف، أو لِنَقُل إنَّها غَيرُ مَعلومةٍ بالضَّرورة لأبناء اللُّغة. ومن ثَمَّ، تستدعي

الـمُعالَجَةُ العميقةُ إجراءاتٍ أوَّليَّة تسبقُ الـمُعالجَةَ الآليَّةَ للمحارف، بهدف استنباط قواعدَ تنبني عليها مُعطَياتٌ جديدة، ثُمَّ تدريب الآلة على استيعاب هذه الـمُعطَيات.

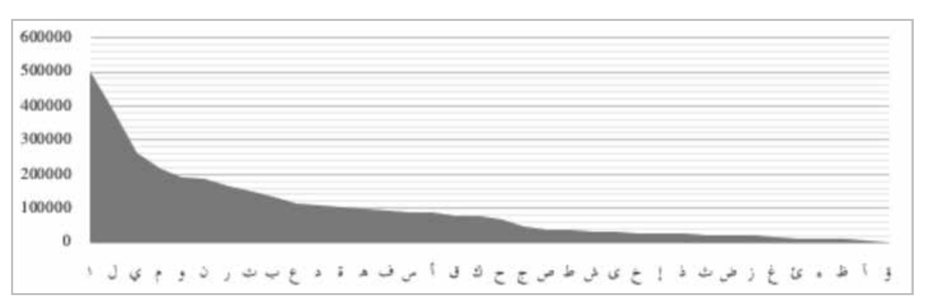
وفي إطار مُعالجَة المحارف العربيَّة، سواءٌ أكانَت مُفرَدةً أم مُرَكَّبة، يُمكنُ التَّمثيلُ على الـمُعالَجَة العميقة بظاهِرَتَين، هما: ائتلاف الحُرُوف، ودوران الحُرُوف. وذلكَ على النَّحو الآتي:

١) ائتلاف الحُـــرُوف [المحارف] العربيَّة.

يُمكنُ توجيهُ الآلة إلى التَّعرُّف على أشكال المحارف العربيَّة ومواضع وُرُودِها في الكلمة، ويُمكنُ توجيهُها كذلكَ إلى الأقسام الصَّرفيَّة (السَّوابق، والجُذُوع، واللَّواحق) لأجل تحليلها كتابيًّا. لكنَّ هُناكَ قواعدَ أخرى لائتلاف الحُرُوف العربيَّة فيما بينَها. ومنها مثلًا: أنَّ حرف العَين لا يأتلفُ مع الغَين، وأنَّ الحاءَ لا يأتلفُ مع الهاء، ومثلُ ذلكَ في حُرُوفٍ أخرى. ويُساعدُ استخلاصُ قواعد ائتلاف الحُرُوف على توجيه الآلة إلى مُعالَجة أخطاء الكتابة العربيَّة.

٢) دَوَران الحُـــــرُوف [المحارف] العربيَّة.

يشتملُ النِّظامُ الكتابيّ في اللُّغة العربيَّة على ثمانيةٍ وعشرينَ حرفًا [جرافيمًا] أساسيًّا؛ وإذا أرَدنا حَصرَ تنويعات هذه الحُرُوف، فسنجدُها ستَّةً وثلاثينَ تنويعًا، تشملُ: سبعةَ تنويعاتٍ للهمزة (ا، آ، أ، إ، ء، ؤ، ئ)، ء، أ،)، وتنويعَين للتَّاء (ت، ة)، وسبعةً وعشرينَ تنويعًا للأحرُف السَّبعة والعشرين الأخرى. لكنَّ مُعالجَةً عميقةً لهذه التَّنويعات تجعلُنا نكتشفُ أنَّ حُرُوفَ العربيَّة وتنويعاتِها ليسَت في ذات الـمُستوى من الدَّوران في النُّصُوص العربيَّة؛ حيثُ تردُ بعضُها بكثرة، مثل (ا، ل، ي، م، ن)، وتردُ بعضُها بصورةٍ مُتوسِّطة، مثل (ق، ك، ح، ج)، ويندُرُ وُرُودُ بعضِها الآخَر مُقارنةً بغَيره، مثل (ء، ظ، آ، ؤ)، على النَّحو الوارد في (الشَّكل ٣). ويُستفادُ من هذه الـمُعطَيات في توجيه الآلة إلى الاحتمالات الغالبة للمحارف العربيَّة حالَ وُرُودِها في النُّصُوص، الأمر الَّذي يُساعدُ على تعيين مواطن الخطإ، واقتراح البدائل الـمُناسبة.



الشَّكل ٣: دوران المحارف [ألوجرافات الحُروف] في النُّصُوص العربيَّة

٣,٢. من تطبيقات مُعالَجة اللُّغة العربيَّة المكتوبة في المُستوى الجرافيميّ.

تحظَى الـمُعالَجَةُ الآليَّةُ لمحارف اللُّغات الإنسانيَّة بعناية العاملينَ في ميادين حَوسَبة اللُّغة؛ نظرًا للحاجة إليها في رَقمَنَة التُّراث الإنسانيّ المكتوب، وإيجاد بيئةٍ تفاعُليَّةٍ لذوي الاحتياجات الخاصَّة من الرَّاغبينَ في قراءة الكُتُب، وتطوير عمل الهواتف الذَّكيَّة، ودواعٍ أخرى. ولا يختلفُ منهجُ مُعالجة الجرافيمات كثيرًا بينَ اللُّغات الإنسانيَّة؛ حيثُ يُعَوَّلُ أساسًا على أشكال المحارف ومواضع الائتلاف بينها. ويُمكنُ التَّمثيلُ على تطبيقات مُعالَجة العربيَّة المكتوبة في الـمُستوى الجرافيميّ بما يأتي.

١,٣,٢. التَّعرُّف الآليّ على الحُرُوف المطبوعة Typewritten OCR.

من تطبيقات التَّعرُّف الآليّ [الضَّوئيّ] على المحارف Optical Character Recognition (OCR)؛ ويُستفادُ من هذا التَّطبيق في تحويل النُّصُوص المطبوعة من صيغتِها الورقيَّة (مثل الكُتُب والوثائق ونحوها) إلى صيغةٍ رقميَّةٍ [مُحَوسَبة]، يُمكنُ التَّحكُّمُ في محارفها وكلماتها بالإضافة أو الحذف أو التَّعديل؛ أو بعبارةٍ أخرى: تحويل الصُّورة إلى نَصٍّ مكتوب. ويستدعي هذا التَّطبيقُ مُراعاةَ التَّبايُن الواقع في أشكال المحارف المطبوعة؛ من حيثُ ألوانُها وأنواعُها وأحجامُها؛ كما يستدعي مُراعاةَ قواعد ائتلاف المحارف، ومواطن وُجود النِّقاط على المنقوط منها، ومواطن وُجود علامات الضَّبط ونحوِها.

٢,٣,٢. التَّعرُّف الآليّ على الحُرُوف المخطوطة Handwriting OCR.

هو أيضًا من تطبيقات التَّعرُّف الآليّ على المحارف OCR؛ ويُستفادُ منهُ في تحويل النُّصُوص المخطوطة [المكتوبة بخطّ اليد] إلى صيغةٍ رقميَّة. ويُساعدُ هذا التَّطبيقُ في مجالاتٍ عديدة، مثل: مُعالجة بصمة الكتابة، وتحقيق النُّصُوص، وتحرير الوثائق المخطوطة. ويستدعي هذا التَّطبيقُ مُراعاةَ التَّبايُن الواقع في أشكال المحارف المخطوطة؛ وهو - بطبيعة الحال - أكثرُ تعقيدًا من سابقه، نظرًا لاختلاف أنماط الكتابة عبرَ الزَّمان والمكان والأفراد، بالإضافة إلى تأثير الضَّوضاء البصريَّة الـمُحيطة بالوثائق المخطوطة.

٢,٣,٣. قراءة الكُتُب Book Reader.

يهدفُ هذا التَّطبيقُ إلى توجيه الآلة إلى القيام بدَور القارئ الآليّ للنُّصُوص في الكُتُب والوثائق المكتوبة؛ حيثُ يُساعدُ ذوي الاحتياجات الخاصَّة من فاقدي القُدرة على الإبصار على القراءة. ويَمُرُّ بمرحَلَتَين رئيسَتَين؛ حيثُ تُعنى المرحلةُ الأولى بالتَّعرُّف الآليّ على المحارف [الجرافيمات]، لتتحوَّلَ المادَّةُ بذلكَ من (نصّ مُصَوَّر جامد) إلى (نَصّ مكتوب قابل للـمُعالَجَة). أمَّا المرحلةُ الأخرى، فتُعنى بتحويل النَّصّ المكتوب إلى نَصٍّ منطوقٍ، عبرَ تطبيقٍ آخَرَ لتحويل النَّصّ إلى كلام Text To Speech (TTS).

٣. المُعالَجَة الآليَّة البنويَّة [الصَّرفيَّة].

٣,١. وَحدة المُعالَجَة الصَّرفيَّة.

تُعنى الـمُعالَجَة البنويَّة [الصَّرفيَّة] بمجموعةٍ من العناصر الـمُكوِّنة لبنية الكلمة في اللُّغات الطَّبيعيَّة، سواءٌ بتحليلها إلى مُكوِّناتٍ صُغرى أم بتوليدها من مُكوِّناتٍ صُغرى. وَوَحدةُ الـمُعالَجَة البنويَّة هي (المورفيم Morpheme)؛ ويُمكنُ تعريفُهُ بأنَّهُ أصغرُ وَحدةٍ لُغويَّةٍ مُجرَّدةٍ لها معنًى. ويكونُ هذا المعنى مُعجميًّا، على نحو ما نجدُ في مورفيمات (إنسان، حُبّ، سماء) أو وظيفيًّا، على نحو ما نجدُ في مورفيمات (الـ) الدَّالَّة على التَّعريف، و(سـ) الدَّالَّة على التَّسويف، و (ـة) الدَّالَّة على التَّأنيث؛ ونحو ذلك.

وتتداخلُ المورفيماتُ في الكلمة الواحدة لتُنتِجَ عناصِرَ بنيتِها الـمُتمثِّلة فيما يأتي:

- الجذر Root: يُقصَدُ به مجموعة الحُرُوف الأصليَّة الـمُكوِّنة للكلمة الـمُشتقَّة أو القابلة للاشتقاق [مثل: الكلمات الـعربيَّة والـمُعرَّبة]. ويُمكنُ أن يُطلَقَ عليه (الأصل) إذا كانَ لكلمةٍ غَير مُشتقَّة [مثل: الكلمات الدَّخيلة والضَّمائر].
- الجِذع Stem: يُقصَدُ به صُورةُ الكلمة الـمُعَيَّنة بعدَ تجريدِها من زوائدِها [السَّوابق واللَّواحق].
- الفرع Lemma: هُوَ الـمُقابلُ الصَّرفيُّ الـمُجرَّدُ للوحدة الـمُعجميَّة [Lexeme] في الكلمة الـمُعَيَّنة. ويأتي في العربيَّة على صُورة الفعل الماضي الـمُسنَد إلى الغائب [ما لم يَلزم صُورةً أخرى]، أو الاسم الـمُفرَد في صيغة التَّذكير [ما لم يَلزم صُورةً أخرى]، أو الأداة في شكلها الرَّئيس.
- الزَّائدة Affix: وهي مُورفيمٌ لاصقٌ، يسبقُ جذعَ الكلمة في بدايته، فيكونُ (سابقة Prefix) أو يلتحقُ به في نهايتِه، فيكونُ (لاحقة Suffix).

ويُمكنُ التَّمثيلُ على هذه العناصر في الكلمة (سيُعاودون) على النَّحو الوارد في (الجدول ٣).

م	العُنصُر		التَّمثيل
١	الجذر		ع و د
٢	الجِذع		يُعاوِد
٣	الفرع		عاوَدَ
٤	الزَّائدة	السَّابقة	سـ
		اللَّاحقة	ـون

الجدول ٣: العناصر الـمُكوِّنة لبنية الكلمة العربيَّة - نموذج (سيُعاودون)

٢,٣. التَّوجيه الآليّ لمُعالجة الوحدات الصَّرفيَّة في اللُّغة العربيَّة.

يُعَدُّ النِّظامُ الصَّرفيُّ العربيُّ مثاليًّا في حَوسَبَتِه، نظرًا لقياسيَّة جُزءٍ كبيرٍ من قواعده، الأمرُ الَّذي يسمحُ بتطويع الآلة للتَّعاطي مع هذه القواعد، بعدَ ترجمتِها من اللُّغة الطَّبيعيَّة إلى لُغات الآلة. ومع هذا، فإنَّ هناكَ جانبًا سماعيًّا ليسَ هيِّنًا في الصَّرف العربيّ. وسنُحاولُ الإبانةَ - فيما يأتي -عن التَّوجيه الآليّ لـمُعالجَة الوحدات الصَّرفيَّة العربيَّة، عبرَ الـمُستويَين: السَّطحيّ والعميق.

١,٢,٣. المُعالَجَة السَّطحيَّة للوحدات الصَّرفيَّة.

يُعنى الصَّرفُ العربيُّ بالأفعال والأسماء دونَ الأدوات الَّتي تَلزمُ حالةَ الجُمود. ويُصَنَّفُ الفعلُ والاسمُ إلى عدَّة أبوابٍ صرفيَّة، وفقًا لاعتباراتٍ مُختلِفة. ويُوضِّحُ (الجدول ٤) بعضًا من هذه الأقسام.

<table>
<tr><th>م</th><th>قِسم الكلام</th><th>مِعيار التَّصنيف</th><th colspan="2">الباب الصَّرفيّ</th><th>التَّمثيل</th></tr>
<tr><td rowspan="3">١</td><td rowspan="7">الفعل</td><td rowspan="3">باعتبار الزَّمَن</td><td colspan="2">الماضي</td><td>كَتَبَ</td></tr>
<tr><td colspan="2">الـمُضارع</td><td>يَكتُبُ</td></tr>
<tr><td colspan="2">الأمر / الطَّلَب</td><td>اكتُب</td></tr>
<tr><td rowspan="2">٢</td><td rowspan="2">باعتبار الصِّحَّة والاعتلال</td><td colspan="2">الصَّحيح</td><td>بَعَثَ</td></tr>
<tr><td colspan="2">الـمُعتلّ</td><td>هَوى</td></tr>
<tr><td rowspan="2">٣</td><td rowspan="2">باعتبار التَّجرُّد والزِّيادة</td><td colspan="2">الـمُجَرَّد</td><td>دَرَسَ</td></tr>
<tr><td colspan="2">المزيد</td><td>دارَسَ</td></tr>
<tr><td rowspan="2">٤</td><td rowspan="8">الاسم</td><td rowspan="2">باعتبار الجُمُود والاشتقاق</td><td colspan="2">الجامِد</td><td>الرَّجُل</td></tr>
<tr><td colspan="2">الـمُشتَقّ</td><td>الـمُستَخدِم</td></tr>
<tr><td rowspan="4">٥</td><td rowspan="4">باعتبار الصِّحَّة والاعتلال</td><td colspan="2">الصَّحيح</td><td>الكِتاب</td></tr>
<tr><td rowspan="3">الـمُعتلّ</td><td>المنقوص</td><td>القاضي</td></tr>
<tr><td>المقصور</td><td>الفتى</td></tr>
<tr><td>الـمَمدود</td><td>الصَّحراء</td></tr>
<tr><td rowspan="2">٦</td><td rowspan="2">باعتبار التَّجرُّد والزِّيادة</td><td colspan="2">الـمُجَرَّد</td><td>السَّهْم</td></tr>
<tr><td colspan="2">المزيد</td><td>الاستِقدام</td></tr>
</table>

الجدول ٤: مِن أبواب الصَّرف العربيّ

وتُصَنَّفُ هذه الأبوابُ بدَورِها إلى أبوابٍ فرعيَّةٍ أخرى. فعلى مُستوى الأفعال – مثلًا – يأتي الفعلُ الـمُجَرَّدُ ثُلاثيًّا ورُباعيًّا؛ ويأتي الثُّلاثيُّ منهُ في ثلاثة أبوابٍ (سماعيَّة) بصيغة الماضي، هي (فَعَلَ، فَعِلَ، فَعُلَ)، ويأتي كذلك في ستَّة أبوابٍ (سماعيَّة) باعتبار صيغتَي الماضي والـمُضارع معًا، هي (فَعَلَ يَفعُل «مثل: نَصَرَ يَنْصُر»، فَعَلَ يَفعِل «مثل: جَلَسَ يَجْلِس»، فَعَلَ يَفعَل «مثل: ذَهَبَ يَذْهَب»، فَعِلَ يَفعَل «مثل: عَلِمَ يَعْلَم»، فَعُلَ يَفعُل «مثل: حَسُنَ يَحْسُن»، فَعِلَ يَفعِل «مثل: حَسِبَ يَحْسِب»)، وهكذا. وعلى مُستوى

الأسماء، يأتي الاسمُ الـمُشتَقُّ من الفعل الثُّلاثيّ في صُورة اسم الفاعل على زنة (فاعِل)، وفي صُورة اسم المفعول على زِنة (مفعول)، ويكونُ قياسيًّا في أكثر الأحيان، وسماعيًّا في مواضعَ قليلة، كأن يُستَغنى عن اسم الفاعل مثلًا مع الفعل (شَجُعَ) لتحلَّ محلَّهُ الصِّفة الـمُشَبَّهة (شُجاع)، وأن يأتي الاسمُ الـمُشتقُّ في صُورة اسم المفعول على زِنة (فَعيل)، مثل (قَتِيل). وهكذا.

وإذا أخذنا في الاعتبار:

- أنَّ التَّمييزَ بينَ الأبواب الصَّرفيَّة يستدعي ضَبطَها بالشَّكل، للتَّمييز مثلًا بينَ اسم الفاعل (مُعَلِّم) واسم المفعول (مُعَلَّم) الـمُشتَقَّيْن من الفعل الرُّباعيّ (عَلَّمَ)،
- وأنَّ الـمُفرَدات في أبواب الصَّرف العربيّ تَقبلُ دُخُولَ الزَّوائد عليها، من السَّوابق واللَّواحق،
- وأنَّ أبوابَ الصَّرف العربيّ الرَّئيسة تتفرَّعُ إلى أبوابٍ أخرى، تُنتِجُ قواعدَ سماعيَّة وقياسيَّة.

فسيكونُ علينا أن نقِفَ على الإشكالات الـمُحتَمَلة حالَ توجيه الآلة إلى الـمُعالَجَة الصَّرفيَّة.

وعلى سبيل المثال، إذا أردنا إجراءَ مُعالَجَةٍ صرفيَّةٍ لكلمة (فلك) بصُورتِها الـمُجَرَّدة من علامات الضَّبط، فستكونُ النَّتيجةُ عدَّةَ احتمالاتٍ، يُصَنِّفُ بعضُها الكلمةَ فعلًا، ويُصنِّفُ بعضُها الكلمةَ اسمًا، ويُصَنِّفُ بعضُها الكلمةَ حرفًا أو أداة. ويُبيِّنُ (الجدول ٥) بعضًا من هذه الاحتمالات.

م	الكلمة	السَّوابق	الجذع	اللَّواحق	قِسم الكلام
١	فَلَكَ	–	فَلَكَ	–	فِعل
٢	فَلَّكَ	–	فَلَّ	ك	فِعل
٣	فُلْك	–	فُلْك	–	اسم
٤	فَلَك	–	فَلَك	–	اسم
٥	فَلَكَ	ف	ل (للجرّ)	ك	حرف / أداة

الجدول ٥: مِن احتمالات الـمُعالَجَة الصَّرفيَّة لكلمة (فلك) بصورتها الـمُجَرَّدة

لأجل مُعالجة مِثل هذه الإشكالات في اللُّغة العربيَّة، فإنَّنا نقومُ بتزويد الآلة بمجموعةٍ من قواعد البيانات الـمُشتملة على مُعطياتٍ لُغويَّة تتبعُ منظومةَ التَّصريف العربيّ (بجانِبَيه: القياسيّ، والسَّماعيّ)؛ وتُمثِّلُ هذه الـمُعطَياتُ أساسًا لتوجيه الآلة إلى مُعالجة الوحدات الصَّرفيَّة، عبرَ خوارزماتٍ مُعَيَّنةٍ مُوافقةٍ لطبيعة العربيَّة وبنيتِها الصَّرفيَّة. ويُوضِّحُ (الجدول ٦) نماذجَ من هذه الـمُعطَيات وأمثلةً عليها.

م	الـمُعطَيات	النَّموذَج	التَّمثيل
١	قوائم الأفعال والأسماء ذات الطَّبيعة القياسيَّة، وقواعدها التَّصريفيَّة والاشتقاقيَّة	الأفعال الرُّباعيَّة والخُماسيَّة والسُّداسيَّة	(ساعَدَ، تأنَّى، استَخدَمَ)
		الأسماء الـمُشتقَّة من الأفعال الرُّباعيَّة والخُماسيَّة والسُّداسيَّة	اسم الفاعل (مُسْتَعِين) من الفعل السُّداسيّ (اسْتعانَ)
٢	قوائم الأفعال والأسماء ذات الطَّبيعة السَّماعيَّة	الأفعال الثُّلاثيَّة الـمُجَرَّدة (بصيغتَي الماضي والـمُضارع)	قَرَأَ - يَقْرَأ كَتَبَ - يَكْتُب
		الأسماء الجامدة	(الحُبّ، العَيْن، الإنسان)
٣	قوائم التَّنويعات الكتابيَّة	تنويعات الكلمات الأعجميَّة	(إنجليزيَّة، إنكليزيَّة، إنغليزيَّة)
٤	قوائم الأدوات	أدوات جَزم الـمُضارع	(إنْ، لَمْ، «لا» النَّاهية)
٥	قوائم حُرُوف الزِّيادة على الأفعال	س، ء، ل، ت، م، و، ن، ي، هـ، ا	التَّاء والألِف في (تَقارَبَ)
٦	قوائم السَّوابق	سوابق العَطف	(و، فـ)
٧	قوائم اللَّواحق	لواحق الجمع والتَّأنيث	(ـون، ـين، ـات، ـة)

الجدول ٦: مِن مُعطَيات الـمُعالجَة السَّطحيَّة للوَحدات الصَّرفيَّة في اللُّغة العربيَّة

تجري مُعالجَةُ الوَحدات الصَّرفيَّة استنادًا إلى هذه الـمُعطَيات؛ ويتحدَّدُ شكلُ الـمُخرَجات بناءً على النُّصُوص الـمُستهدَفة؛ حيثُ تختلفُ مُخرَجاتُ الكلمات

الـمشكولة كُلِّيًّا عن الكلمات الـمشكولة جُزئيًّا، وتختلفُ مُخرَجاتُ الكلمات المشكولة عن مُخرَجات الكلمات الـمُجرَّدة من الضَّبط. وبطبيعة الحال، فإنَّ احتمالات الـمُعالَجة الصَّرفيَّة تقلُّ إلى الحدّ الأدنى إذا كانت الكلمةُ تامَّةَ التَّشكيل، وتزدادُ تدريجيًّا كُلَّما قلَّ الضَّبط، حتَّى تصلَ إلى الحدّ الأقصى في الكلمات الـمُجرَّدة من الضَّبط.

٢,٢,٣. المُعالَجة العميـــقـــة للوحدات الصَّرفيَّة.

تتَّسمُ اللُّغةُ العربيَّةُ بوُجود ظاهرة الإعراب. ومع هذا، فواقعُ الكتابة العربيَّة أن تخلُوَ الكلماتُ من علامات الضَّبط، إلَّا في مواضع الالتباس أو بعض أنماط النُّصُوص الـمُقَدَّسة أو التُّراثيَّة أو نحو ذلك. ومن ثَمَّ، فالنَّتيجةُ الطَّبيعيَّةُ لـمُعالَجة الوحدات الصَّرفيَّة أن تتعدَّدَ احتمالاتُها إلى الحدّ الأقصى، لأنَّ الآلَةَ تُقدِّمُ كُلَّ احتمالٍ مُمكنٍ للكلمة الـمُجرَّدة من الضَّبط حالَ ضبطِها بمُختلَف الأشكال الـمُمكنة. وعلى سبيل المثال، ستتعدَّدُ احتمالاتُ مُعالَجة كلمة (حمل) بصُورتِها الـمُجَرَّدة بينَ الأسماء والأفعال؛ وإذا حوَّلناها إلى صُورةٍ مضبوطةٍ جُزئيًّا بوضع السُّكون في وسط الكلمة على هيئة (حمْل)، فهذا يعني استثناءَ الأفعال؛ وإذا حوَّلناها إلى صُورةٍ مضبوطةٍ كُلِّيًّا [على مُستوى البِنية] على هيئة (حِمْل)، فهذا يعني الاقتصارَ على نتيجةٍ واحدةٍ [أو نتائجَ محدودة الاحتمالات]؛ وهكذا.

وثمَّةَ نُقطةٌ مُهمَّةٌ أخرى تَستَرعي الانتباهَ بشأن مُوافَقة مُخرَجات الـمُعالَجة الصَّرفيَّة للُّغة العربيَّة الـمُستخدَمة. فالواقعُ أنَّ عددًا ليسَ هيِّنًا من الـمُخرَجات تقعُ في دائرة الصَّحيح؛ لكنَّها لا تُوجدُ في الواقع؛ وإذا وُجِدَت فهي نادرةٌ للغاية. ويُمكنُ التَّمثيلُ على ذلك بالوَحدات الصَّرفيَّة في جُملة (هذا هو أخي). فاستخدامُ الكلمات الثَّلاثة يشيعُ في العربيَّة على صُورة الضَّبط (هَذا هُوَ أخي) بدلالة (هَذا) على اسم الإشارة، و (هُوَ) على ضَمير الغائب، و (أخي) على الاسم الـمُسنَد إلى ياء الـمُتكلِّم؛ إلَّا أنَّ مُخرَجات الـمُعالَجة الصَّرفيَّة ستُشيرُ إلى احتمال أن تكونَ (هذا) فعلًا ماضيًا مُسنَدًا إلى ألف الاثنَين (هَذَّا)، وأن تكونَ (هُوَ) فعلًا للطَّلب (هَوِّ)، وأن تكونَ (أخي) كلمةً مُصَغَّرَة (أُخَيّ)؛ ونحو ذلك. ومثلُ هذه الاحتمالات صحيحٌ من النَّاحية المعياريَّة؛ لكنَّهُ نادرُ الوُجود جدًّا في واقع اللُّغة.

زُبدَةُ القَول أنّنا ننشُدُ في الـمُعالَجَة العميقةِ للوحدات الصَّرفيَّة أمرَين، هُما:

١) توجيه الآلة إلى دلالات الكلمات الـمُجرَّدة من الضَّبط.

٢) ترتيب احتمالات مُعالَجَة الوَحدات الصَّرفيَّة بحسب دورانِها في واقع اللُّغة.

ويستدعي الأمران تدريبَ مُدوَّنةٍ لُغويَّةٍ على الـمُعالَجَة العميقة؛ حيثُ تُعَدُّ هذه الـمُدوَّنةُ موردًا لـمُعطَيات القرائن الـمُساعدة على ترجيح الدِّلالات الأقرب إلى الصَّواب؛ مثل قرينة أن تكونَ كلمة (من) اسمًا موصولًا إذا لَحِقَتها كلماتٌ تبدأُ بحروف (ت، ن، ي)؛ حيثُ تُرَجَّحُ إشارةُ هذه الكلمات إلى أفعالٍ مُضارعة، أو نحو ذلك. وبطبيعة الحال فإنَّ دورانَ الـمُفرَدات في الـمُدوَّنة يُعَدُّ مُوَجِّهًا للآلة لترتيب الاحتمالات وتنقيتها من مُخرَجاتٍ لا تُعَبِّرُ عن واقع اللُّغة، وإن كانت خاضعةً للقياس.

٣,٣. من تطبيقات مُعالَجَة اللُّغة العربيَّة المكتوبة في المُستوى الصَّرفيّ.

ما دُمنا نتعاملُ مع مُفرَدات النُّصُوص [الأبنية]، فالحاجةُ ماسَّةٌ ومُستمرَّة إلى حَوسَبة الصَّرف العربيّ في مُختلَف ميادين حوسَبة اللُّغة، بطريقٍ مُباشرٍ أو غَير مُباشِر. وتَبرُزُ الحاجةُ إلى الـمُعالَجَة الآليَّة لقواعد الصَّرف العربيّ في العديد من التَّطبيقات عبرَ عمليَّتَين أساسيَّتَين، هما (التَّوليد) الَّذي ينطلقُ من الجذر [الأصل] اللُّغويّ بهدف تعيين الوحدات [المورفيمات] وتشكُّلاتها الصَّرفيَّة، و (التَّحليل/ التَّفكيك) الَّذي ينطلقُ من الكلمة [المجموع الكتابيّ] بهدف تعيين الـمُكوِّنات الصَّرفيَّة وأصولها الَّتي تكوَّنَت عنها. ويُمكنُ التَّمثيلُ على تطبيقات مُعالَجَة العربيَّة المكتوبة في الـمُستوى الصَّرفيّ بما يأتي.

١,٣,٣. الفهرسة الآليَّة Automatic indexing.

تُعنى الفهرسةُ الآليَّةُ ببناء قوائم الكلمات وتتابُعاتِها من مجموعات النُّصُوص، وترتيبها على نَسَقٍ مُعَيَّنٍ، كأن تُرَتَّبَ وفقًا لألفبائيَّة اللُّغة، أو بحسب دوران الكلمات في النَّصّ، أو نحو ذلك. وتتأثَّرُ مُخرَجاتُ أدوات الفهرسة الآليَّة باختلاف النّظام الصَّرفيّ للُّغة الطَّبيعيَّة؛ حيثُ تبدو في صورة وحداتٍ مُتَّسقة مع البنية الـمُعجميَّة في اللُّغات الإلصاقيَّة (مثل: الإنجليزيَّة، والفرنسيَّة)، وتبدو أقلَّ اتِّساقًا في اللُّغات الاشتقاقيَّة (مثل: العربيَّة، والعبريَّة). والإشكالُ بالنِّسبة للُّغة العربيَّة أنَّ أدوات الفهرسة الألفبائيَّة

تتأثَّرُ بزوائد الكلمة وتنويعاتها الاشتقاقيَّة، فتقومُ بتوزيعها على مداخلَ عديدة. وسعيًا إلى حلّ هذا الإشكال، فإنَّنا نلجأُ إلى بناء ما يُعرَفُ بـ (أدوات الفهرسة الجذريَّة) الَّتي تقومُ على مُعالجَةٍ صرفيَّةٍ للكلمات بهدف تجاوز الزَّوائد وتجميع التَّنويعات الاشتقاقيَّة.

٢,٣,٣. التَّحليل الصَّرفيّ Morphological Analysis.

تُؤدِّي أدواتُ التَّحليل الصَّرفيّ دورًا مُهمًّا في مُعالجَة النَّصّ العربيّ؛ ويُعَوَّلُ عليها بصورةٍ كبيرةٍ في حَوسَبة الـمُعجم العربيّ، لا سيَّما فيما يتَّصلُ بمبانيه [الوحدات والمداخل الـمُعجميَّة]. وتتفاوتُ مُخرَجاتُ التَّحليل الصَّرفيّ بحسب وسيلة الـمُعالجَة؛ إذ يُمكنُ الاقتصارُ على تعيين الوحدات الصَّرفيَّة وتوصيفاتها؛ ويُمكنُ تجاوزُها إلى عمليَّاتٍ أخرى، تشملُ: (التَّجذير Rooting) الَّذي يُعنى باستخلاص جُذُور الكلمات أو أصولها اللُّغويَّة، و (التَّجذيع Stemming) الَّذي يُعنى باستخلاص جُذُوع الكلمات بعدَ تجريدِها من الزَّوائد [السَّوابق واللَّواحق]، و (التَّفريع Lemmatization) الَّذي يُعنى بتعيين ما يُعرَفُ بـ «الفُرُوع»؛ وهي في العربيَّة صُورةُ الوحدات الـمُعجميَّة الـمُنبثقة عن الوحدات الصَّرفيَّة. ومن ناحيةٍ أخرى، تتعدَّدُ التَّنويعاتُ الشَّكليَّة للكلمات بحسب الضَّبط بالشَّكل؛ ويُعتَمَدُ في إدارتها على خوارزماتٍ مُعَيَّنة.

٤. المُعالجَة الآليَّة التَّركيبيَّة.

٤,١. وَحدة المُعالجَة التَّركيبيَّة.

إذا كانت الـمُعالجَة البنويَّة مَعنيَّةً بالـمُفرَدات ذاتِها، من حيثُ بنيتُها التَّصريفيَّة، فإنَّ الـمُعالجَةَ التَّركيبيَّةَ تُعنى بمواقع هذه الـمُفرَدات في التَّركيب النَّحويّ، وتُعنى كذلكَ بالعلاقات الشَّكليَّة بينَ هذه الـمُفرَدات. ويُمكنُ القولُ - بعبارةٍ أخرى - إنَّ الـمُعالجَةَ التَّركيبيَّة ترتكزُ على (أقسام الكلام Parts of Speech) حينَ تقعُ في كيانٍ تركيبيٍّ مُكتمل الأركان؛ سواءٌ أكانَ هذا الكيانُ بسيطًا أم مُركَّبًا.

في ضوء ذلكَ، لا نستطيعُ الوُقُوفَ على وَحدةٍ ثابتةٍ للـمُعالجَة التَّركيبيَّة، نظرًا لتبايُن الكيانات الَّتي تردُ فيها أقسامُ الكلام في اللُّغات الطَّبيعيَّة. فقد تكونُ هذه الوَحدةُ (عِبارة Phrase) أو (جُملة Sentence) أو (فِقرة Clause)، وفقًا لطبيعة التَّركيب الـمُتضمِّن

لأقسام الكلام وموضعه في السِّياق. وتوضيحًا لذلك، يُمكنُ التَّمثيلُ على الوحدات التَّركيبيَّة في العربيَّة على النَّحو الوارد في (الجدول ٧).

م	الكيان التَّركيبيّ	نماذج الوَحدة التَّركيبيَّة
١	العبارة	الأُمم الـمُتَّحدة
٢	الجُملة	الأُمم الـمُتَّحدةُ مُنظَّمةٌ دوليَّةٌ فعَّالة
٣	الفِقرة	يُمكنُ اللُّجوءُ إلى الأُمم الـمُتَّحدة؛ حيثُ تُعَدُّ مُنظَّمةً دوليَّةً فعَّالة

الجدول ٧: أنماط الوحدات التَّركيبيَّة في اللُّغة العربيَّة [العبارة، الجُملة، الفِقرة]

سنُلاحظُ من خلال هذه النَّماذج ما يأتي:

- النَّمَط (١) – العبارة: تدلُّ العبارة (الأُمم الـمُتَّحدة) على كيانٍ تركيبيٍّ، حيِّزُهُ أقلُّ من الجُملة؛ لكنَّ هذا الكيانَ يُمثِّلُ وحدةً تركيبيَّةً كاملةً حينَ يردُ في سياق عُنوانٍ أو لافتةٍ أو نحو ذلك.
- النَّمَط (٢) – الجُملة: تدلُّ الجُملة (الأُمم الـمُتَّحدةُ مُنظَّمةٌ دوليَّةٌ فعَّالة) على كيانٍ تركيبيٍّ، حيِّزُهُ الجُملة الكاملة؛ حيثُ يُعبِّرُ عن جُملةٍ اسميَّةٍ رُكناها الـمُبتدأ والخَبر. ويُمثِّلُ وحدةً تركيبيَّةً حينَ يردُ في سياقٍ لُغويٍّ مُكتمل.
- النَّمَط (٣) – الفِقرة: تدلُّ الفِقرة (يُمكنُ اللُّجوءُ إلى الأُمم الـمُتَّحدة؛ حيثُ تُعَدُّ مُنظَّمةً دوليَّةً فعَّالة) على كيانٍ تركيبيٍّ، حيِّزُهُ أكبر من الجُملة [في هذه الفقرة جُملتان]. ويُمثِّلُ هذا الكيانُ وحدةً تركيبيَّةً حينَ يردُ في سياق مجموعةٍ من الجُمَل الـمُترابطة الَّتي تدورُ حولَ فكرةٍ مُعيَّنة، ويُعَوَّلُ على إحداها في فهم تركيب غَيرِها.

٢, ٤ . التَّوجيه الآليّ لمُعالجة الوحدات التَّركيبيَّة في اللُّغة العربيَّة.

حينَ نتعاطى مع الجُملة باعتبارِها وَحدةَ الـمُعالَجة التَّركيبيَّة، فإنَّنا نُميِّزُ في العربيَّة بينَ نَوعَيها: الفعليَّة والاسميَّة. ويتكوَّنُ النَّوعُ الأوَّلُ (الجُملة الفعليَّة) من عُنصُرَين رئيسَين – في حالة اللُّزُوم – هما: الفعل والفاعل (أو الفعل ونائب الفاعل)، ومن ثلاثة عناصِرَ – في حالة التَّعدِّي، هي: الفعل والفاعل والمفعول. أمَّا الجُملةُ الاسميَّة، فلها

عُنصُران رئيسان، هُما: الـمُبتدأ والخَبَر. وثمَّةَ عناصرُ أخرى ثانويَّة، تُؤدِّي وظائفَ الإضافة والعطف والوَصف والتَّوكيد وغَير ذلك.

وحينَ نُوجِّهُ الآلة إلى فهم طبيعة هذه العناصِر، فعلَينا أوَّلًا أن نُحدِّدَ أقسامَ الكلام الَّتي تنتمي إليها عناصرُ الجُملة؛ وعلينا تاليًا أن نُوجِدَ قرائنَ [أو علامات] مُميِّزةٍ لكُلِّ عُنصُرٍ على حِدة؛ بحيثُ تكونُ هذه العلاماتُ مِقياسًا يُستَنَدُ إليه في تمييز العُنصُر الواحد عن العناصِر الأخرى. ويبدو الأمرُ بسيطًا للوَهلة الأولى، إذ يُفتَرَضُ أنَّ لكلِّ قسمٍ من أقسام الكلام (الاسم، والفعل، والحرف/ الأداة) علاماتٍ مُميِّزة وضَعَها النُّحاةُ عندَ التَّقعيد للعربيَّة؛ كالتَّعريف والتَّنوين لتمييز الاسم، وتاء الفاعل ونُون التَّوكيد لتمييز الفعل، والتَّجرُّد من سِمات الاسم والفعل لتمييز الحَرف.

لكنَّ الأمرَ لا يخلو من إشكالاتٍ غَير قليلة حينَ نُوجِّهُ الآلةَ للـمُعالجَة التَّركيبيَّة؛ وعلى سبيل المثال: إذا وَجَّهنا الآلةَ إلى اعتبار الكلمات المسبوقة بـ (ال) أسماءً، فالنَّتيجةُ لن تكونَ صحيحةً في مثل الأفعال (التَحمَ، التزَم، التَقى، التَمَسَ، ...)؛ وإذا وَجَّهنا الآلةَ إلى اعتبار الكلمات الَّتي تَلحَقُ بها التَّاء (ــت) أفعالًا، فالنَّتيجةُ لن تكونَ صحيحةً في مثل الأسماء (بَيْت، سَبْت، نَبْت، نَعْت، ...)؛ وقِس على ذلكَ مُختَلَف مُوجِّهات القواعد التَّركيبيَّة للُّغة العربيَّة. ومن ثَمَّ، فالقواعدُ المفهومة، ممَّا يستطيعُ العقلُ البشريُّ تمييزُه، قد لا تكونُ مفهومةً للآلة على الوجه الأمثل. وسنُحاولُ أن نستبينَ ذلكَ فيما يأتي:

١,٢,٤. المُعالجَة السَّطحيَّة للوحدات التَّركيبيَّة.

في هذا الـمُستوى من الـمُعالجَة، يكونُ الهدفُ تَعيينَ أقسامِ الكلام. والواقعُ أنَّ للُّغة العربيَّة تصنيفاتٍ مُتعدِّدةً لأقسام الكلام، منها التَّصنيفُ الثُّلاثيُّ التَّقليديّ (الاسم، والفعل، والحرف)، والتَّصنيف الرُّباعيّ (الاسم، والفعل، والحرف، والضَّمير)، والخُماسيّ (الاسم، والفعل، والحرف، والضَّمير، والظَّرف)، والسُّباعيّ (الاسم، والفعل، والحرف، والضَّمير، والظَّرف، والصِّفة، والخالفة). وأيًّا كانَ التَّصنيفُ الَّذي نعتمدُهُ في توجيه الآلة، فالمنهجُ واحدٌ؛ إذ يقومُ على تدريب الآلة على مُدوَّناتٍ لُغويَّة مُوَسَّمة Annotated Corpora بعلاماتٍ مُعَيَّنةٍ يُستَدَلُّ بها على أقسام الكلام PoS.

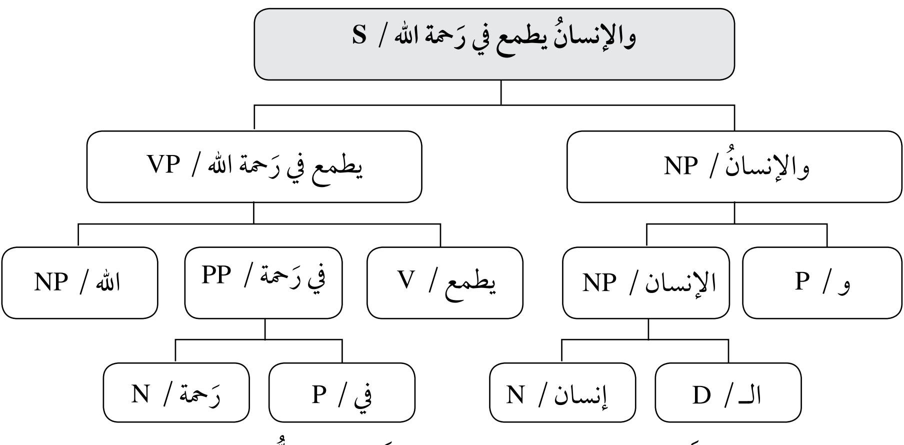
ولتوضيح ذلكَ يُمكنُ القولُ إنَّ الـمُعالَجَةَ السَّطحيَّةَ للوحدات التَّركيبيَّة تتمُّ عبرَ عمليَّةٍ، تُعرَفُ بـ «العَنونة التَّركيبيَّة Syntactic Annotation». ويتفرَّعُ عن هذه العمليَّة إجراءان، على النَّحو الآتي:

١) **التَّوسيم Tagging**: يتمُّ على مُدوَّنةٍ لُغويَّةٍ كبيرةٍ نسبيًّا؛ ويُقصَدُ به إلحاقُ علاماتٍ مُعيَّنة، تُعرَفُ بـ «الوُسُوم التَّركيبيَّة Syntactic Tags» بالكلمات الـمُتَضَمَّنة في هذه الـمُدوَّنة، لتتحوَّلَ بذلك من مُدوَّنةٍ خام Raw Corpus إلى مُدوَّنة مُوَسَّمة Annotated Corpus. وليسَت للوُسُوم التَّركيبيَّة صورةٌ ثابتة، سواءٌ من حيثُ الكمُّ أم الكيف؛ إذ تتحكَّمُ فيها أهدافُ الـمُعالَجَة. ويُمكنُ التَّمثيلُ عليها بالنَّموذج الوارد في (الشَّكل ٤).

النَّصُّ قبلَ التَّوسيم
الآن .. كيفَ يُمكنُ أن نُحصِيَ نُجومَ السَّماء ؟
النَّصُّ بعدَ التَّوسيم
الآن[AD]/ كيفَ[QU]/ يُمكنُ[VI]/ أن[PO]/ نُحصِيَ[VI]/ نُجومَ[CN]/ السَّماء[CN]/ ؟/
وُسُوم أقسام الكلام PoS Tags
الظَّرف [AD]، الاستفهام [QU]، الفعل المضارع [VI]، الأداة [PO]، الاسم الشَّائع [CN]

الشَّكل ٤: نموذج لمادَّة نصِّيَّة مُستَمدَّة من مُدوَّنة لُغويَّة مُوَسَّمة تركيبيًّا

٢) **التَّجـــزئة Parsing**: ويتَّصلُ هذا الإجراءُ بالوَحدة التَّركيبيَّة في صُورتِها الـمُكتملة؛ حيثُ نقومُ من خلالِهِ بتقسيم هذه الوحدة على مُكوِّناتٍ أصغَرَ، لنتمكَّنَ من التَّعامُل مع كُلٍّ منها على حِدة. ويُمكنُ التَّمثيلُ على هذا الإجراء بالنَّموذج الوارد في (الشَّكل ٥).



الشَّكل ٥: نموذج لتجزئة الوَحدة التَّركيبيَّة في اللُّغة العربيَّة

يُعَوَّلُ على مُخرَجات العَنونة التَّركيبيَّة - الَّتي تتمُّ في مرحلة ما قبلَ الـمُعالجَة - في توفير المورد الرَّئيس للـمُعالجَة التَّركيبيَّة؛ ونعني (الـمُدوَّنة اللُّغويَّة الـمُوَسَّمة تركيبيًّا Syntactic Annotated Corpus). ونقومُ - عَبرَ هذا المورد - بتدريب النُّصُوص، لنخلُصَ إلى قاعدة بياناتٍ لـمُفرَدات اللُّغة ووُسومِها الـمُمثِّلة لأقسام الكلام. وعبرَ عمليَّات الـمُطابقة، يُمكنُ للآلة أن تُحاكي مُخرَجات قاعدة البيانات، حالَ توجيهها إلى الـمُعالجَة التَّركيبيَّة السَّطحيَّة بتعيين أقسام الكلام.

٢,٢,٤. المُعالجَة العميـــــقـــة للوحدات التَّركيبيَّة.

لا تكفي القاعدةُ التَّركيبيَّةُ الصِّرفة لتوجيه الآلة إلى مُعالجةٍ سليمة؛ إذ تُحاطُ القاعدةُ في كثيرٍ من الأحيان ببعض الاستثناءات. وقد يتخلَّلُها التباسٌ ناتجٌ عن تشابُه بعض أقسام الكلام أو تجرُّدها من علامات الضَّبط [التَّشكيل] الـمُميِّزة لها. ومثالُ ذلك ما نجدُهُ في كلمة (بل) الَّتي يصعبُ تحديدُ قسمِها الكلاميّ في صُورتِها الـمُجرَّدة؛ حيثُ تحتملُ أن تكونَ اسمًا (بَلٌّ) أو فعلًا (بَلَّ) أو حرفًا (بَلْ)؛ ومثلُ ذلكَ كثيرٌ في العربيَّة. لهذا، تستدعي مُعالجةُ العميقة للوحدات التَّركيبيَّة تجاوُزَ البنية السَّطحيَّة للجُملة إلى تكوين هيكلٍ من القواعد الـمُنتظمة الَّتي تخضعُ لمقياسٍ كمِّيٍّ، يُمكنُ الاستنادُ إليه في توجيه الآلة إلى مُعالجَةٍ مُوافقةٍ لطبيعة اللُّغة. وبطبيعة الحال، فإنَّ هذه القواعدَ تتوافقُ في جُزءٍ منها مع منطق اللُّغة؛ لكنَّها تخرُجُ كذلكَ في أحايينَ كثيرةٍ عن هذا المنطق [غَير الـمُطَّرِد] إلى منطق الآلة [الـمُطَّرِد].

إنَّنا نعتمدُ - في الـمُعالَجة العميقة للوحدات التَّركيبيَّة - على ثلاثة عناصِرَ أساسيَّة، هي:

١) القواعد النَّحويَّة Grammar: تُمَثِّل الـمُعطَيات اللُّغويَّة؛ ومَصدرُها في العربيَّة قوانينُ النَّحو العربيّ [في الجانب التَّركيبيّ منه]؛ لا سيَّما تلكَ الَّتي تتَّصلُ بأقسام الكلام والعلاقات بينها، وأنماط الإعراب والبناء [الرَّفع، والنَّصب، والخَفض، والجزم] لأقسام الكلام الـمُختلفة.

٢) الـمُدوَّنات اللُّغَويَّة المشكولة والـمُوَسَّمة تركيبيًّا: مَصدرُها مجموعاتٌ من النُّصُوص الـمُمثِّلة لواقع اللُّغة؛ ويُفتَرَضُ أن تكونَ هذه الـمُدوَّنات مِقياسًا للاستعمال اللُّغويّ. ووجود هذا العُنصُر في الـمُعالجة التَّركيبيَّة العميقة يعني أنَّ إنجازَها على نحوٍ سليمٍ يستدعي أن تُسبَقَ بمُعالَجةٍ سطحيَّة. وننشُدُ في هذه الـمُدوَّنات أن تُضبَطَ باستخدام علامات الضَّبط العربيَّة Arabic Diacritics؛ حيثُ يُعَوَّلُ عليها في تعيين المواقع الإعرابيَّة لأقسام الكلام [بأنماطِها الـمُعرَبة والمبنيَّة]؛ ويُعوَّلُ عليها كذلكَ في الاستدلال على الأنماط الإعرابيَّة للكلمات الـمُتَضَمَّنة في الوَحدة التَّركيبيَّة الـمُعَيَّنة، سواءٌ أكانَت هذه الكلماتُ مُلازمةً لحالة الإعراب أم البناء.

٣) خوارزميَّة التَّحليل التَّركيبيّ Syntactic Analysis Algorithm: وهي الـمُتسلسلة الرِّياضيَّة الَّتي تُعبِّرُ عن مراحل الـمُعالَجة التَّركيبيَّة. ويُعَوَّلُ فيها على الطُّرُق الإحصائيَّة الَّتي تُساعدُ على حَصر أنماط الجُملة وتوصيفِها على نحوٍ مُوافقٍ للُّغة الـمُستخدَمة فعليًّا.

تتآلَفُ هذه العناصرُ فيما بينها لتكوين منظومة الـمُعالَجة التَّركيبيَّة في اللُّغة العربيَّة. فالقواعدُ النَّحويَّةُ هي الأساسُ الَّذي يتحدَّدُ من خلالِهِ شكلُ الوَحدة التَّركيبيَّة؛ إذ إنَّها الـمُوجِّهُ الرَّئيسُ لمعرفة أركان الجُملة [الأساسيَّة والثَّانويَّة]، وقواعد ضبطِها على نَسَقٍ سليم. والـمُدوَّنةُ الـمُوَسَّمة هي المورد الَّذي نستبينُ من خلالِه أنماطَ الجُملة العربيَّة ونعرفُ حُدُودَها وأساليبَ صَوغِها في اللُّغة الحيَّة. وخوارزميَّةُ التَّحليل التَّركيبيّ هي الوسيلةُ القياسيَّةُ الَّتي تقودُنا إلى أنماط التَّراكيب الشَّائعة وغَير الشَّائعة، وتقودُنا كذلكَ إلى تعيين احتمالات الـمُعالَجة الأقرب إلى واقع اللُّغة حالَ وُجُود التباسٍ في النُّصُوص.

ولمزيدٍ من التَّوضيح، نعرضُ في (الجدول ٨) نماذجَ مُوجِّهات العناصِر الثَّلاثة، مع التَّمثيل.

<table>
<tr><th>م</th><th>العُنصُر</th><th>الـمُوَجِّه</th><th colspan="2">التَّمثيل</th></tr>
<tr><td>١</td><td>القواعد النَّحويَّة</td><td>القاعدة ١: «لن» أداة مَبنيَّة، تسبق الفعل الـمُضارع، وتنصِبُه.</td><td colspan="2">لن [أداة] تكونَ [فعل مُضارع منصوب].</td></tr>
<tr><td>٢</td><td>الـمُدوَّنات اللُّغويَّة</td><td>القاعدة ٢: كُلُّ ما بدأ بـ«يسـ» يُوسَمُ بأنَّهُ (فعل مُضارع).</td><td colspan="2">يَسْبَح - يَسْتَشْمِر</td></tr>
<tr><td rowspan="3">٣</td><td rowspan="3">خوارزميَّة التَّحليل التَّركيبيّ</td><td>في القاعدة ١: القاعدة صحيحة، مع وُجُود استثناءاتٍ يُظهرُها الضَّبط.</td><td colspan="2">لِنْ [فِعل طَلَب] لأخيكَ</td></tr>
<tr><td rowspan="2">في القاعدة ٢: القاعدة صحيحة، مع وُجُود استثناءاتٍ يُظهرُها النَّصّ.</td><td>استثناء مُؤكَّـــد:</td><td>يَسُوع</td></tr>
<tr><td>استثناء مُحتَمَل:</td><td>يَسِير</td></tr>
</table>

الجدول ٨: نماذج من مُوجِّهات عناصر الـمُعالَجَة التَّركيبيَّة العميقة في اللُّغة العربيَّة

سنُلاحظُ أنَّ القاعدةَ النَّحويَّةَ تُعبِّرُ عن أحد القوانين التَّركيبيَّة في العربيَّة؛ لكنَّ توجيهَ الآلة إلى هذا القانون يفرضُ تزويدَها بمُعطياتٍ استثنائيَّة، لتتعاطى مع الاحتمالات الأخرى. وسنُلاحظُ كذلكَ أنَّ القاعدَةَ الـمُستَخلَصة من الـمُدوَّنة اللُّغويَّة تُعبِّرُ عن أحد القوانين الوصفيَّة الَّتي يفرضُها النَّصّ. ونقومُ في هذه الحالة أيضًا بتزويد الآلة بمُعطياتٍ استثنائيَّة [مُؤَكَّدة ومُحتَمَلة]. وفي القاعدتَين [النَّحويَّة والنَّصِّيَّة]، يُعَوَّلُ على خوارزميَّة التَّحليل التَّركيبيّ في تتبُّع الاستثناءات وتقدير احتماليَّة وُرُودِها.

٣, ٤. من تطبيقات مُعالَجَة اللُّغة العربيَّة المكتوبة في المُستوى التَّركيبيّ.

يُمثِّلُ مُستوى الـمُعالَجَة التَّركيبيَّة تحدِّيًا للعاملينَ في حوسَبة اللُّغة، نظرًا لطبيعة الجُملة العربيَّة وتعدُّد أنماطِها وثراء العربيَّة بالظَّواهر التَّركيبيَّة ذات الطَّبيعة الخاصَّة [كالأسماء الخمسة، والتَّثنية، والجُموع السَّالمة] ووُجود ظاهرة الإعراب [الَّتي تتحكَّمُ في شكل الكلمات في حالَتَي البناء والإعراب] ونحو ذلك. ويُمكنُ التَّمثيلُ على تطبيقات مُعالَجَة العربيَّة المكتوبة في الـمُستوى التَّركيبيّ بما يأتي.

٤,٣,١. التَّدقيق الإملائيّ Spell and Grammar Checking.

يُعنى التَّدقيقُ الإملائيُّ باستكشاف أخطاء الكتابة؛ سواءٌ أكانَت في قواعد إملاء الكلمات أم في القواعد التَّركيبيَّة (النَّحويَّة)؛ ويَعقُبُ استكشافَ الأخطاء تقديمُ اقتراحٍ لمجموعةٍ من احتمالات الصَّواب. ويُعَدُّ هذا التَّطبيقُ أحدَ أبرز تطبيقات الـمُعالجَة التَّركيبيَّة في العربيَّة، نظرًا للحاجة إليه مِن قِبَل مُستَخدِمي العربيَّة في التَّفاعُل مع الآلة بأشكالها الـمُختلفة [الحواسيب اللَّوحيَّة، والهواتف الذَّكيَّة، ...].

ويستمدُّ هذا التَّطبيقُ مُعطَياتِهِ من مُدوَّنةٍ لُغويَّةٍ مُتعدِّدة الأنماط التَّركيبيَّة، على أن تكونَ نُصُوصُها مكتوبةً بصورةٍ سليمة [إملائيًّا ونحويًّا]؛ حيثُ يتمُّ تدريبُ الـمُدوَّنة على مُطابقة الـمُفرَدات والتَّراكيب الـمُستَهدَفة بالـمُفرَدات والتَّراكيب الـمُتَضَمَّنة في قواعد بيانات أدوات التَّدقيق، ثُمَّ اقتراح احتمالات الصَّواب عندَ اكتشاف وُجُود خطإٍ ما، استنادًا إلى مُعجمٍ [أو مَسرَد] يحوي قدرًا كبيرًا من الـمُفرَدات والـمُتلازمات؛ حيثُ تُرَجَّحُ الكلماتُ الصَّحيحة الَّتي تشتركُ مع الكلمة موضع الخطإ في عددٍ من المحارف أو الَّتي يكثُرُ تلازمُها مع كلماتٍ سابقةٍ أو لاحقة بالكلمة الـمَعنيَّة.

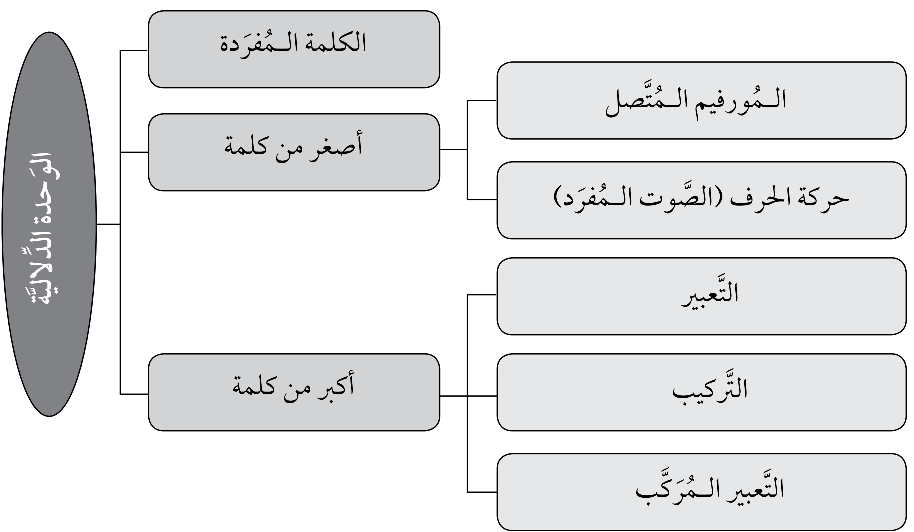
٤,٣,٢. التَّشكيل الآلي Automatic Diacritics.

يفرضُ النِّظامُ التَّركيبيُّ للُّغة العربيَّة وُجُودَ مثل هذا التَّطبيق الَّذي يُعنى بضبط شكل الكتابة باستخدام (علامات الضَّبط). ويعملُ هذا التَّطبيقُ في العربيَّة عبرَ مُستوَيَين؛ يُعنى الأوَّلُ منهما بضبط الأبنية [الكلمات الـمُجَرَّدة]، ويُعنى الآخَر بضبط الإعراب [أواخر الكلمات]. ويستمدُّ هذا التَّطبيقُ مُعطَياتِهِ من مُدوَّنةٍ لُغويَّةٍ مُوَسَّمةٍ ومَضبوطةٍ بالشَّكل ضبطًا تامًّا؛ ويعتمدُ في ضبط الأبنية بصورةٍ كبيرةٍ على دَوانِ شَكل الكلمة في الـمُدوَّنة الـمُوَجَّهة للتَّدريب على الضَّبط؛ إذ تزدادُ احتماليَّةُ الضَّبط بشكلٍ مُعَيَّنٍ بازدياد دوران الكلمة بهذا الشَّكل. أمَّا في مُستوى الإعراب، فيعتمدُ على قرائنَ لُغويَّةٍ مُعَيَّنةٍ تُصاحبُ الكلمات العربيَّة (مثل: أدوات نَصب الأفعال، وحُرُوف جرّ الأسماء، ونحوها)، بالإضافة إلى اعتمادِهِ على مُخرَجات إحصاء تدريب أنماط الجُمَل [البسيطة والـمُرَكَّبة] في الـمُدوَّنة اللُّغويَّة.

٥. المُعالَجة الآليَّة الدِّلاليَّة.

١,٥. وَحدة المُعالَجة الدِّلاليَّة.

وَحدَةُ الـمُعالَجَة الدِّلاليَّة هي (السِّيميم Sememe)؛ وهو أصغرُ وَحدةٍ دلاليَّةٍ لها معنًى. ونظرًا لتوزُّع مُفرَدات اللُّغة بينَ (الكلمات التَّامَّة Full Words) الَّتي يُعنى بها الـمُعجمُ، كالأسماء والأفعال والصِّفات، و (الكلمات الصُّوريَّة Form Words) الَّتي يُعنى بها علمُ التَّركيب [النَّحو]، كالضَّمائر والأدوات؛ فإنَّ الوَحدةَ الدِّلاليَّةَ تأخذُ أحدَ ثلاث صُوَر، على النَّحو الـمُوَضَّح في (الشَّكل ٦). ويُمكنُ التَّمثيلُ على صُوَر الوحدات الدِّلاليَّة بالنَّماذج الـمُوَضَّحة في (الجدول ٩).



الشَّكل ٦: أنماط الوَحدة الدِّلاليَّة في اللُّغات الطَّبيعيَّة

م	أنماط الوحدات		نماذج الوَحدة الدِّلاليَّة
١	الكلمة الـمُفرَدة (الـمُورفيم الـمُفرَد)		إنسان
٢	أصغر من كلمة	الـمُورفيم الـمُتَّصل	(سـ) ـيعمل = التَّسويف
		حركة الحرف	كَتَبـ (ـتَ) = خطاب الـمُفرَد الـمُذكَّر
٣	أكبر من كلمة	التَّعبير	ضَرَبَ أخماسًا في أسداس = تحيَّرَ
		التَّركيب	تأبَّطَ شرًّا = العلميَّة
		التَّعبير الـمُركَّب	الحافلة النَّهريَّة

الجدول ٩: أنماط الوحدات الدِّلاليَّة في اللُّغة العربيَّة

٢,٥. التَّوجيه الآليّ لمُعالجة الوحدات الدِّلاليَّة في اللُّغة العربيَّة.

إذا جاءَت الوَحدةُ الدِّلاليَّة على نَمَطٍ أصغَر من الكلمة الواحدة، فإنَّها تُشيرُ إلى معنًى وظيفيّ Functional Meaning؛ وتُوَجَّهُ آليًّا حينئذٍ للـمُعالجَة الصَّرفيَّة أو التَّركيبيَّة؛ وإذا جاءَت على نَمَطٍ أكبرَ من الكلمة، فإنَّها تُشيرُ إلى معنًى مُصطلحيّ Terminological Meaning، يُهتَدى إليه عبرَ قواعد بياناتٍ خاصَّة، تحوي قوائمَ الـمُصطلحات والتَّعابير والتَّراكيب الـمُتلازمة ومعانيها؛ وتكونُ هذه القواعدُ أشبَهَ بـ(المعاجم الـمُصطلحيَّة). أمَّا إذا جاءَت الوَحدةُ الدِّلاليَّة على نَمَط الكلمة الـمُفرَدَة، فإنَّها تُشيرُ إلى معنًى مُعجميّ Lexical Meaning؛ وهُوَ الَّذي تتعاطى معهُ الآلةُ عندَ توجيهِها إلى الـمُعالجَة الدِّلاليَّة. ويتمُّ هذا التَّوجيهُ عبرَ الـمُستويَين: السَّطحيّ والعميق على النَّحو الآتي.

١,٢,٥. المُعالجَة السَّطحيَّة للوحدات الدِّلاليَّة.

إذا بحَثنا في (الـمُعجم الوسيط) عن كلمة (البداوة)، فسنجدُ المعنى الآتي:

(البَداوة) الحياة في البادِية؛ ويَغلُبُ عليها التَّنقُّل والتِّرحال.

وإذا بحثنا في الـمُعجم ذاتِهِ عن كلمة (العَين)، فسنَجدُ المعاني الآتية:

(العَيْن) عُضْو الإبصار للْإِنْسَان وَغَيره من الْحَيَوَان،
و: ينبوع المَاء يَنْبع من الأَرْض وَيجْرِي،
و: أهل البَلَد،
و: أهل الدَّار،
و: الجاسوس،
و: رَئِيس الجَيْش،
و: ...

فالوَحدةُ الدِّلاليَّةُ تَحتملُ أن تُشيرَ إلى معنًى واحدٍ، وتحملُ كذلكَ أن تُشيرَ إلى أكثرَ من معنًى. وفي الحالة الأولى (عندَ الدِّلالة على معنًى واحد)، يُمكنُ توجيهُ الآلة إلى هذا المعنى عَبرَ (الـمُعجَم Dictionary) باعتباره موردًا رئيسًا للمعاني الـمُعجميَّة؛ مع مُراعاة أنَّ

بعضَ الكلمات تُشيرُ إلى معانٍ مجازيَّة ليسَت مُتَضَمَّنةً بالضَّرورة في الـمُعجَم. أمَّا في الحالة الأخرى (عندَ الدِّلالة على أكثر من معنى)، فعلَينا حينئذٍ أن نُوجِّهَ الآلة إلى هذه المعاني جميعًا عبرَ مُعالَجَة سَطحيَّةٍ تستندُ إلى قواعد بياناتٍ مُعجميَّة؛ وسيكونُ الإشكالُ في هذه الحالة في توجيهِ الآلة إلى ترجيح المعنى المقصود من جُملة المعاني الـمُقترَحة.

والواقعُ أنَّ المعاني الـمُعجميَّة لا تكفي وحدَها لـمُعالَجة الوحدات الدِّلاليَّة؛ إذ يُعَوَّلُ كذلكَ على العلاقات القائمة بينَ الوَحدات الدِّلاليَّة. ويُبيِّنُ (الجدول ١٠) بعضًا من هذه العلاقات.

م	العلاقة		المفهوم	التَّمثيل
١	أُفقيَّة	التَّرادُف	إذا كانَ (أ) يُعبِّر عن (ب)، و (ب) يُعبِّر عن (أ)، فإنَّ (أ) و (ب) مُترادفان.	الخَوف = الخَشية
٢		التَّضادّ	إذا كانَ (أ) ضدّ (ب)، و (ب) ضدّ (أ)، فإنَّ (أ) و (ب) مُتضادَّان.	الأخذ × العطاء
٣		التَّنافُر	إذا كانَ (أ) يشتمل على (ب) و (ج)، وكانَ (ب) ≠ (ج)، فإنَّ (ب) و (ج) مُتنافران.	الرَّجُل ≠ المرأة (إنسان)
٤	رأسيَّة	النَّوعيَّة	إذا كانَ (أ) نوعًا من (ب)، فالعلاقة بينَ (أ) و (ب) نوعيَّة.	التِّين – الفاكهة
٥		الجُزئيَّة	إذا كانَ (أ) جُزءًا من (ب)، فالعلاقة بينَ (أ) و (ب) جُزئيَّة.	الرَّأس – الجَسَد
٦		التَّلازُميَّة	إذا كانَ (أ) يُؤدِّي إلى (ب)، وكان (ب) لا يتمُّ إلَّا عن طريق (أ)، فإنَّ (أ) و (ب) مُتلازمان.	اللُّثغة / الكلام

الجدول ١٠: من العلاقات الدِّلاليَّة في اللُّغة العربيَّة

تقومُ الـمُعالَجَةُ السَّطحيَّة للوحدات الدِّلاليَّة على توظيف مواردَ لُغويَّة في توجيه الآلة إلى معاني هذه الوَحدات والعلاقات بينها. وبالإضافة إلى (الـمُعجم)، فثمَّةَ مواردُ أخرى، منها على سبيل المثال:

- الشَّبكة الدِّلاليَّة Semantic Net: وتُعَدُّ هيكلًا معرفيًّا يربطُ بينَ الوحدات [الكلمات] باستخدام العلاقات الدِّلاليَّة في مُستوياتٍ أفقيَّةٍ ورأسيَّة مُتعدِّدة.

- شبكة الكلمات WordNet: وتُعَدُّ قاعدةَ بياناتٍ مُعجميَّة دلاليَّة؛ حيثُ تربطُ بينَ الوحدات والمفاهيم من ناحية، وبينَ الوحدات والعلاقات الدِّلاليَّة من ناحيةٍ أخرى.
- الأنطولوجيا Ontology: وهي شبكةٌ معرفيَّةٌ ضَخمة، تجمعُ في كيانِها بينَ الشَّبكات الدِّلاليَّة وشبكات الكلمات؛ وتتجاوزُ الأنطولوجيا المعاني الـمُعجميَّة إلى مُختَلَف أشكال المعرفة؛ كما تتجاوزُ العلاقات الدِّلاليَّة الـمُباشرة إلى العلاقات غَير الـمُباشرة بينَ الوحدات.

٢, ٢, ٥. المُعالَجَة العميـــــقـــة للوحدات الدِّلاليَّة.

إذا كانَت الـمُعالَجَةُ السَّطحيَّةُ للوحدات الدِّلاليَّة تُعنى بتوجيه الآلة إلى المعنى الواحد أو المعاني الـمُتعدِّدة، فالـمُعالَجَة العميقةُ تُعنى بالبحث في احتماليَّة وجود معانٍ أخرى للوحدات ذات المعنى الواحد، وتُعنى كذلكَ بالتَّرجيح بينَ احتمالات المعاني في الوحدات الَّتي تُشيرث إلى معانٍ مُتعدِّدة. وتتمُّ الـــمُعالَجَةُ العميقةُ عبرَ ما يُعرَفُ بـ «فكّ الالتباس الدِّلاليّ Word Sense Disambiguation»؛ وهو إجراءٌ مُتطوِّرٌ يهدفُ إلى تعيين المعنى المقصود، اعتمادًا على قرائنَ لُغويَّة تُستَخلَصُ من السِّياق الَّذي تردُ فيه الوَحداتُ الدِّلاليَّة. وتقومُ فكرةُ هذا الإجراء على البحث عن المعنى المقصود في مُتجاورات الوَحدة الـمُعَيَّنة؛ حيثُ يغلُبُ أن تتجاورَ الوَحدة الدِّلاليَّة الدَّالَّة على معنًى مُعَيَّنٍ مع أنماطٍ مُعَيَّنةٍ من مُفردات اللُّغة؛ سواءٌ أكانَ هذا التَّجاورُ مُباشَرًا (قبلَ الوَحدة أو بعدَها) أم غَيرَ مُباشر.

في اللُّغة العربيَّة، يُمكنُ التَّمثيلُ على التَّجاوُر الـمُباشِر بكلمة (ابن) حينَ تأتي مُجاورةً [سابقة] لوَحدةٍ دلاليَّة مُعَيَّنة. فإذا رَمَزنا مثلًا للكلمة (ابن) بالرَّمز (S) وللوَحدة الـمُعجميَّة الـمَعنيَّة [التَّالية] بالرَّمز (N)، فإنَّ وُرُودَهما مُتجاورَتَيْن يعني ترجيحَ احتماليَّة أن تُشيرَ الوَحدة (N) إلى العلَميَّة. ومن ناحيةٍ أخرى، يُمكنُ التَّمثيلُ على التَّجاوُر غَير الـمُباشِر بكلمة (يُغمِض) حينَ تأتي في سياق الوَحدة الدِّلاليَّة (عَين) في الجُملة (حينَ يُغمِضُ الإنسانُ عَينَهُ يشعرُ بشيءٍ من السَّكينة)؛ حيثُ أفادَت الكلمة (يُغمِض) الَّتي جاءَت في صُورة الفعل الـمُضارع دلالةَ الوَحدة (عَين) على (عُضو الإبصار)؛ وهكذا.

ونظرًا لطبيعة اللُّغة العربيَّة في أبنيتِها وتراكيبها، فإنَّ فكَّ الالتباس الدِّلاليّ لوَحداتِها يستدعي تداخُلًا بينَ الأنظمة الثَّلاثة (البنويّ، والتَّركيبيّ، والدِّلاليّ) - إلى حدٍّ كبير؛ حيثُ يُعَوَّلُ على النِّظام البنويّ في تعيين الوَحدة الدِّلاليَّة ذاتِها، وتجريدها من زوائدها (السَّوابق واللَّواحق)، ويُعَوَّلُ على النِّظام التَّركيبيّ في تعيين قِسم الكلام الَّذي تنتمي إليه الوَحدة، واستكشاف مُتجاوراتها، وتعيين حُدُود السِّياق اللُّغَويّ؛ ويُعَوَّلُ على النِّظام الدِّلاليّ في تعيين معاني الوَحدة، ثُمَّ ترجيح المعنى الأقرب إليها.

بَقِيَ أن نُشيرَ إلى وجود العديد من خوارزمات فكّ الالتباس الدِّلاليّ WSD Algorithms. ومنها مثلًا: (خوارزك Lesk، خوارزم بايز Naïve Bayesian، خوارزم Yarowsky). والواقعُ أنَّ طرائقَ تطبيق هذه الخوارزمات مُتباينة إلى حدٍّ ما، وإن كانت جميعًا تنطلقُ من قرينة السِّياق اللُّغويّ. وفي العُموم، يستدعي فكُّ الالتباس الدِّلاليّ في العربيَّة تلاحُمًا بينَ مجموعةٍ من الموارد الَّتي تشملُ: (المعاجم والموسوعات، وشبكات الكلمات، والأنطولوجيات، والـمُدوَّنات اللُّغَويَّة، والـمُحلِّلات التَّركيبيَّة).

٣,٥. من تطبيقات مُعالجَة اللُّغة العربيَّة المكتوبة في المُستوى الدِّلاليّ.

يُعَوَّلُ على التَّحليل الدِّلاليّ في مُعالجَة اللُّغة العربيَّة المكتوبة في العديد من التَّطبيقات الَّتي تتَّصلُ بالمعنى؛ مع التَّأكيد على التَّمازُج الحتميّ بينَ مُستوى التَّحليل الدِّلاليّ ومُستويات التَّحليل اللُّغويّ الأخرى، وفقًا للغاية من الـمُعالجَة. ويُمكنُ التَّمثيلُ على هذه التَّطبيقات بما يأتي:

١,٣,٥. استرجاع المعلومات النَّصِّيَّة Text Information Retrieval:

يهدفُ هذا التَّطبيق إلى البحث أو الاستدلال عن معلوماتٍ مُعَيَّنةٍ في مُحتوًى كبيرٍ نسبيًّا من النُّصُوص غَير الـمُنتظمة. ويُستفادُ منهُ في بناء مُحرِّكات البحث Search Engines وهيكلة مُستودعات البيانات Data-Warehouses. وتبرُزُ أهمِّيَّةُ الـمُعالجَة الدِّلاليَّة في هذا التَّطبيق عندَما تحملُ كلمةُ البحث الـمُستهدَفة دلالاتٍ مُتعدِّدة، ويُرادُ التَّركيزُ على نتائج البحث عن إحدى هذه الدِّلالات، على نحو ما نجدُ مثلًا عندَ البحث عن كلمة (الأسد) حالَ قَصْد توجيه البحث إلى العَلَميَّة.

٢,٣,٥. تحليل المشاعر (التَّعرُّف على الآراء) Opinion Mining:

من تطبيقات التَّنقيب في النُّصُوص Text Mining؛ ويهدفُ إلى الكشف عن المشاعر أو الآراء الَّتي تُعبِّرُ عن وُجهات نظر الأفراد والجماعات [تلميحًا أو تصريحًا] في مجموعةٍ من النُّصُوص. ويُستفادُ من هذا التَّطبيق في مجالاتٍ عديدة، مثل: البحث الجنائيّ، والاستكشاف الـمُخابَراتي، والتَّسويق، بالإضافة إلى الدِّراسات المسحيَّة المعنيَّة بتحليل الآراء. ولهذا التَّطبيق أهمِّيَّةٌ كبيرةٌ في تحليل البيانات النَّصِّيَّة في وسائل التَّواصُل الاجتماعيّ Social Media؛ نظرًا لثراء مادَّة هذه الوسائل بالنُّصُوص المكتوبة من ناحية، وتعدُّد توجُّهات مُستخدميها وآرائِهم من ناحيةٍ أخرى.

٣,٣,٥. التَّعرُّف على أسماء الكائنات Named-entity Recognition:

يُعَدُّ كذلكَ أحدَ تطبيقات التَّنقيب في النُّصُوص؛ ويهدفُ هذا التَّطبيقُ إلى تمييز الأسماء الدَّالَّة على العلَميَّة [الكائنات أو الكيانات]، كأسماء الأعلام والمعالم والـمُؤسَّسات ونحوها؛ وهي أسماءٌ يكثرُ وُجُودُها في البيانات النَّصِّيَّة الـمُنتظمة [أو شبه الـمُنتظمة]، كالـمُدوَّنات اللُّغويَّة والمعاجم والموسوعات ونحوِها. وفي اللُّغة العربيَّة يُستفادُ من هذا التَّطبيق في جوانبَ عديدةٍ، مثل: التَّحليل التَّركيبيّ، وتحليل الخطاب، وبناء قواعد بيانات الأعلام، بالإضافة إلى جوانبَ أخرى تتبعُ مَيدانَ التَّنقيب في النُّصُوص، مثل: التَّلخيص الآليّ للنُّصُوص Automatic Text Summarization الَّذي يُعنى بإعادة بناء النَّصّ في صُورةٍ مُختَصَرة، والتَّقييم الآليّ Automatic Text Scoring الَّذي يُعنى بقياس صحَّة النَّصّ.

ببليوجرافيا مرجعيَّة.

1. Abdel-Fattah, Y. (2018). *Arabic Corpus Linguistic*. Edinburgh University Press.
2. Arulmozi, S.; Dash, N. (2018). *History, Features, and Typology of Language Corpora*. Springer.
3. Blokdyk, G. (2018). *Text Mining Complete Self-Assessment Guide*. Emereo Pty Limited.
4. Brezina, V. (2018). *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge University Press.
5. Collins, L. (2019). *Corpus Linguistics for Online Communication: A Guide for Research*. Routledge.
6. Dash, N. S., Ramamoorthy, L. (2018). *Utility and Application of Language Corpora*. Springer.
7. Deng, L.; Liu, Y. (2018). *Deep Learning in Natural Language Processing*. Springer.
8. Ertel, W. (2018). *Introduction to Artificial Intelligence*. Springer.
9. Farr, F.; Murray, L. (2016): *The Routledge Handbook of Language Learning and Technology*. Routledge.
10. Feldman, R.; Sanger, J. (2006): *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. CUP.
11. Gomez, P. C. (2013): *Statistical Methods in Language and Linguistic Research*. I^{sd}.
12. Handford, M. (2018). *Corpus Linguistics for Discourse Analysis: A Guide for Research*. Routledge.
13. Haynes, D. (2018). *Metadata for Information Management and Retrieval: Understanding metadata and its use. Facet Publishing*.
14. Jo, T. (2018). *Text Mining: Concepts, Implementation, and Big Data Challenge*. Springer.

15. Karwowski, W. (2019). *Intelligent Human Systems Integration 2019*. Springer.
16. Kulkarni, A.; Shivananda, A. (2019). *Natural Language Processing Recipes: Unlocking Text Data with Machine Learning and Deep Learning using Python*. Apress.
17. MacIntyre, J.; Maglogiannis, I.; Iliadis, L.; Pimenidis, E. (2019). Artificial Intelligence Applications and Innovations: AIAI 2019 IFIP WG 12.5 International Workshops: MHDW and 5G-PINE 2019, Hersonissos, Crete, Greece, May 24–26, 2019, Proceedings. Springer.
18. Marr, B. (2019). *Artificial Intelligence in Practice: How 50 Successful Companies Used AI and Machine Learning to Solve Problems*. John Wiley & Sons.
19. McEnery, T. (2014). *Arabic Corpus Linguistics. Edinburgh* University Press.
20. McEnery, T.; Meurers, D.; Rebuschat, P. (2017). Experim*ental, Corpus-based and Computational Approaches to Language Learning: Evidence and Interpretation*. Wiley.
21. Mel’čuk, I. (2015). *Semantics: From meaning to text*. John Benjamins Publishing Company.
22. Neamat El, G.; Yee, S. (2018). *Computational Linguistics, Speech and Image Processing for Arabic Language*. World Scientific.
23. Shaalan, k.; Hassanien, A.; Tolba, F. (2017). *Intelligent Natural Language Processing: Trends and Applications*. Springer.
24. Sharp, B.; Sedes, F.; Lubaszewski, W. (2017). *Cognitive Approach to Natural Language Processing*. Elsevier.
25. Sidorov, G. (2019). *Syntactic N-grams in Computational Linguistics*. Springer.

26. Soudi, A.; Bosch, A.; Neumann, G. (2007). *Arabic Computational Morphology: Knowledge-based and Empirical Methods*. Springer Science & Business Media.
27. Srinivasa-Desikan, V. (2018). *Natural Language Processing and Computational Linguistics: A practical guide to text analysis with Python*, Gensim, spaCy, and Keras. Packt Publishing.
28. Tran, T. X. (2019). *Artificial Intelligence: Learning Artificial Intelligence*. Thanh Tran.
29. Walker, B.; Mcintyre, D. (2019). *Corpus Stylistics: Theory and Practice*. Edinburgh University Press.
30. Zizka, J.; Darena, F.; Svoboda, A. (2019). *Text Mining with Machine Learning*. Taylor & Francis Group.

الفصل الثَّالث

المُعالَجة الآليَّة للُّغة العربيَّة المنطوقة

د. أحمد راغب أحمد

ملخص

مع الطفرة التي أحدثتها التقنية في عالمنا المعاصر، وتزامنًا مع الثورة الإلكترونية التي صبغت كل مجالات الحياة، أصبحت تقنيات معالجة الصوت البشري مطلبًا مهمًّا، ليس فقط لتسهيل عمليات التواصل بين المتحدثين، وإنما لأغراض اقتصادية وتجارية وسياسية لا حصر لها، بالإضافة إلى الأبعاد الأكاديمية والاجتماعية التي تتمثل في مساعدة ذوي الاحتياجات الخاصة عن طريق تحويل الزخم العلمي والثقافي المكتوب إلى مادة مسموعة يفيد منها ذوو الاحتياجات الخاصة والأسوياء على حد سواء.

وقد حاول هذا الفصل التعرض للمبادئ التطبيقية التي يتبعها المختصون في مجال تحليل الصوت اللغوي وتوليده، وذلك من خلال إلقاء الضوء على خصائص مكونات الصوت اللغوي التي قُمت بتقسيمها إلى مجموعتين، تتناول المجموعة الأولى منها الصفات الأكوستية التي ترسم ملامح هذا الصوت بغض النظر عن قائله، في حين تتناول المجموعة الأخرى الصفات الذاتية الشخصية التي ترتبط ارتباطًا وثيقًا بمُنْتِج الصَّوْت، والتي تتفاوت خصائصها حسب الفئات الناطقة ذكورًا وإناثًا، أطفالًا أو شبابًا أو شيوخًا، بل إنَّ هذه الاختلافات تلقي بضوئها على خصائص الصوت حسب اختلاف الأفراد في الفئة الواحدة، وهو ما يترتب عليه وجود صفات أو خصائص صوتية لكل فرد تميزه.

وقد اختتمت الفصل بشرح تطبيقي لعملية إنتاج الصوت العربي حاسوبيًّا من خلال عرض لتقنية تحويل النص العربي المكتوب إلى صوت منطوق (Text To Speech)، وعرضٍ آخرَ لتقنية التعرف على الكلام المنطوق (Automatic Speech Recognition).

الكلمات المفتاحية:

التحليل الصوتي الحاسوبي، مكونات الصوت اللغوي، الناطق الآلي، التماثل الأدائي، التباين الأدائي.

١ . الصَّوت اللُّغويّ وعلم الأصوات.

الكلام هو أكثر وسائل التواصل الطبيعية والبديهية وأكثرها تفضيلًا لدى البشر، وقد أورد ابن جني أنَّ حدَّ اللغة أصوت يعبر بها كل قوم عن أغراضهم (ابن جني، ١٩١٣: ٣٣)، وعليه فإن الكلام المنطوق ما هو إلا تعبير عن الأفكار والمشاعر من خلال إصدار ذبذبات مسموعة تنتقل عبر الجهاز الصوتي لدى الإنسان، ليتم استقبالها بواسطة الجهاز السمعي ومن ثم تتم معالجتها دلاليًّا وفق العرف اللغوي الكامن في أذهان الناطقين باللغة.

وقد عرَّف «روبن» الصوت بأنه اضطراب مادي في الهواء يتمثل في قوة أو ضعف ثم في ضعف تدريجي ينتهي إلى نقطة الزوال النهائي (العطية، ١٩٨٣: ٦).

وينشأ الصوت البشري كما ذهب إبراهيم أنيس من ذبذبات مصدرها عند الإنسان الحنجرة، فعند اندفاع النَّفَس من الرئتين يمر بالحنجرة فتحدث تلك الاهتزازات التي بعد صدورها من الأنف أو الفم تنتقل خلال الهواء الخارجي على شكل موجات حتى تصل إلى الأذن (أنيس، ١٩٦١: ٧).

وبوجه عام فإن الصوت هو الأثر السمعي الذي به ذبذبة مستمرة مطردة حتى ولو لم يكن مصدره الجهاز الصوتي (حسَّان، ١٩٩٠: ٥٩).

فالصوت إذن عبارة عن ذبذبات ناتجة عن قوة تنتقل عبر الهواء، وهو كذلك عرض يخرج من النفس مستطيلا متصلا حتى يعرض له في الحلق والفم والشفتين مقاطع تثنيه عن امتداد واستطالة فيسمى المقطع أينما عرض له حرف (ابن جني، ١٩٥٤: ١٨٦).

ويري غساني أنَّ الصوت ظاهرة طبيعية وشكل من أشكال الطاقة، وهو يستلزم وجود جسم في حالة اهتزاز أو تذبذب، وهذه الاهتزازات أو الذبذبات تنتقل عبر وسط معين حتى تصل إلى أذن الإنسان (حميداني، ٢٠١٠: ١٤).

أما إذا انتقلنا إلى علم الأصوات فيمكن تعريفه بأنه علم لغوي في المقام الأول يُعْنَى بدراسة أصوات اللغة حيث ينظر هذا العلم في الأصوات في حد ذاتها من حيث إخراجها، بل وحتى من حيث سماعها، لكن بعض اللغويين يطلقونه ويريدون

بـــه دراســـة التغييـــرات والتحـــولات التـــي تحـــدث فـــي أصـــوات اللغـــة نتيجـــة تطورها (البهنساوي، ٢٠٠٤: ١١).

٢. مصطلحات أساسية.

قبل الولوج إلى مسائل تحليل الصوت اللغوي حاسوبيًّا ينبغي أن تتم الإشارة إلى أهم المصطلحات التي تؤطّر لهذا النوع من الدراسات، وأول هذه المصطلحات وأكثرها ذيوعًا مصطلح تقنيات اللغات البشرية (Human language Technology)، وهو مجال علمي وتطبيقي حديث يجمع بين دراسة اللغة من ناحية وتطبيقات التحليل الحاسوبي من ناحية أخرى، وتهدف تقنيات اللغات البشرية التي يتم الإشارة إليها اختصارًا بـ (HLT) إلى دراسة المناهج التي تتبعها البرامج الحاسوبية والأجهزة الإلكترونية لتحليل النصوص اللغوية والكلام البشري، كما تدرس طريقة إنتاج تلك النصوص والأصوات وتحليلها أو تعديلها والاستحابة لها بطريقة آلية، وتعتمد هذه التقنيات على مجالين فرعيين، أولهما: معالجة اللغات الطبيعية (Natural Language Processing) أو ما يعرف اختصارًا بـ (NLP) وهو حقل فرعي لعلوم الكمبيوتر وهندسة المعلومات والذكاء الاصطناعي المهتم بالتفاعلات بين أجهزة الكمبيوتر واللغات البشرية (الطبيعية)، وخاصة كيفية برمجة أجهزة الكمبيوتر لمعالجة وتحليل كميات كبيرة من بيانات اللغة الطبيعية. وثانيهما مجال اللُّغويَّات الحاسوبيَّة (Computational Linguistics) ويعرف اختصارًا بـ (CL). مع الاستفادة من معطيات مجالات مكملة أخرى مثل: تقنيات الكلام البشري (Speech Technology) وعلم الإحصاء التطبيقي (Applied Statistics) (Uszkoreit, ٢٠١٨).

وإلى جانب ما سبق أضحت حقول رياضية أخرى - وثيقة الصلة بهيكلة اللغة حاسوبيًّا، مثل نظرية المخططات وبالأخص علم الإحصاء (Statistics) - تحتل مكانة تزداد أهميتها يوما بعد يوم داخل خريطة اللغويات الحاسوبية وتمثيلها الصوري وطبيعتها الديناميكية مما يؤدي إلى فهم آلي أعمق وتحليل أدق للغات الطبيعية (wataonline.net).

وإذا كانت الدراسات اللغوية الحاسوبية العامة قد قطعت شوطًا كبيرًا على مستوى الإنجازات والنتائج فإنَّ الدراسات اللسانية العربية أيضًا قد طرحت عدة محاولات

تهدف إلى تطويع تقنيات الحاسوب للغة العربية بما يتوافق مع شخصيتها ومحارفها ورسومها من جهة، وبما يتواءم مع قواعد اللغة العربية وخصائصها الحاسوبية من جهة أخرى، بدءًا ببرمجة الحروف والنصوص، ومرورًا ببناء التقنيات الصوتية اللغوية لهدف تحسين الاتصال الآلي بين الإنسان والحاسوب (الوعر، ١٩٨٩: ٢٣).

٣. التحليل الحاسوبي لمكونات الصوت اللغوي.

إذا أردنا أن نقدم تحليلًا حاسوبيًّا لمكونات الصوت اللغوي فينبغي أن نجزِّء هذه المكونات إلى مجموعتين، تتناول المجموعة الأولى الصفات الثابتة للصوت المنطوق، وأقصد بها تلك الصفات الأكوستية التي ترسم ملامح هذا الصوت بغض النظر عن قائله، وقد اصطلحت على تسميتها بالتماثل الأدائي، في حين تتناول المجموعة الأخرى الصفات الذاتية الشخصية التي ترتبط ارتباطًا وثيقًا بمُنْتِجِ الصَّوْت، وهو هنا المتكلم، وتتفاوت خصائصها حسب الفئات الناطقة ذكورًا وإناثًا، أطفالًا وشبابًا وشيوخًا، بل إنَّ هذه الاختلافات تلقي بضوئها على خصائص الصوت حسب اختلاف الأفراد في الفئة الواحدة، وهو ما يترتب عليه وجود صفات أو خصائص صوتية لكل فرد تميزه عن غيره من الناطقين الذين ينتمون إلى نفس المجموعة الصوتية، وهذا الأمر يؤدي بدوره إلى وجود بصمة صوتية تميز كل متحدث عن الآخر، وقد اصطلحت على تسمية هذه المجموعة بالتباين الأدائي.

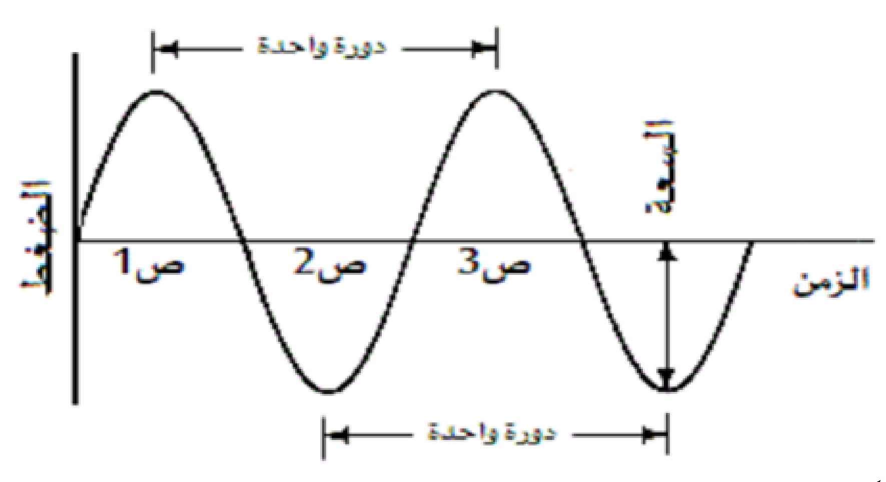
١,٣. التماثل الأدائي:

المراد بالتماثل الأدائي تلك الخواص الصوتية الأكوستية (Acoustics features) التي تتصف بالثبات وتبين مواصفات الصوت اللغوي بغض النظر عن قائله، وبالتالي فلا نجد اختلافًا ملحوظًا في هذه المكونات مرتبطًا بشخصية الناطق أو المصوِّت باعتبار نوعه أو جنسه أو عمره، وعليه فهي أقرب ما تكون إلى الخواص الإستاتيكية، حيث إنها أكثر عمومية وأقل خصوصية، وتتحدد ملامح هذا التماثل الصوتي في الموجة الصوتية (Sound wave)، وسرعة الصوت (Speed of sound)، وشدة الصوت (Sound intensity)، والرَّسم التَّذبذبيُّ (شكل الموجة)، والصورة الطيفية (Spectrogram).

١, ١, ٣. الموجة الصوتية (Sound wave):

الموجة الصوت هي نمط الاضطراب الناتج عن حركة الطاقة التي تنتقل عبر وسيط (وسط) مادي - مثل الهواء أو الماء أو أي مادة سائلة أو صلبة أخرى - حيث تنتشر بعيدًا عن المصوِّت أو مصدر إنتاج الصوت، ونعني به في سياقنا هذا الناطق البشري أو الآلي؛ حيث يُصدر هذا الـمُصَوِّتُ طاقة تُسبِّب اهتزازًا ينتقل من المصدر إلى أقصى نقطة سمعية عبر الوسط المادي، وَيُحْدِثُ هذا الاهتزاز زعزعة لاستقرار جزيئات هذا الوسط المحيط بشكل جزئي؛ ثم تُحْدِثُ هذه الجزئيات إزعاجًا مماثلًا للجزيئات التي تليها وهلم جرا إلى نهاية المدى الصوتي. ومن ثم تنشأ الدوائر الصوتية التي تشبه وتماثل الدوائر التي تحدث عن إسقاط كتلة مادية في مياه البحر، وتحمل تلك الموجاتُ الطاقةَ الصوتيةَ عبر الوسيط في كل الاتجاهات، وتقل كثافتها كلما بعدت عن مصدرها.

وعليه فإنَّ الإشارة إلى الموجة الصوتية تعبر عن إحدى أشكال انتقال الطاقة الصوتية، والتي تنتقل عبر أي وسط مادي يسمح بانتقال الطاقة من المصدر إلى حيز الاستقبال بدون إزاحة جزيئات هذا الوسط بشكل دائم، بل يقتصر الانتقال على الشكل الجزئي، أي أنه لا تنتقل أي كتلة مع انتقال الموجة، ولكن جزيئات الوسط تتحرك بشكل متعامد أو موازٍ لاتجاه حركة الموجة حول موقع ثابت. وتتصف أي موجة بصفة الدورية، وهو ما يترتب عليه أن تكون تكرارا لنمطٍ ما من الشدة في فترات زمنية متتابعة، ويسمى هذا التكرار بالتردد، وهو عبارة عن عدد الموجات المارة في مقطع ما مقسومًا على وحدة الزمن. بينما تسمى المسافة الأفقية التي تقطعها الموجة الواحدة «طول الموجة»، وهو أصغر جزء متكرر مكون للموجات، ويساوي المسافة بين قمتين أو قاعين متتابعين، أو هو المسافة المحصورة بين ثلاث قيم صفرية متتابعة للموجة، كما يوضحه الشكل التالي.

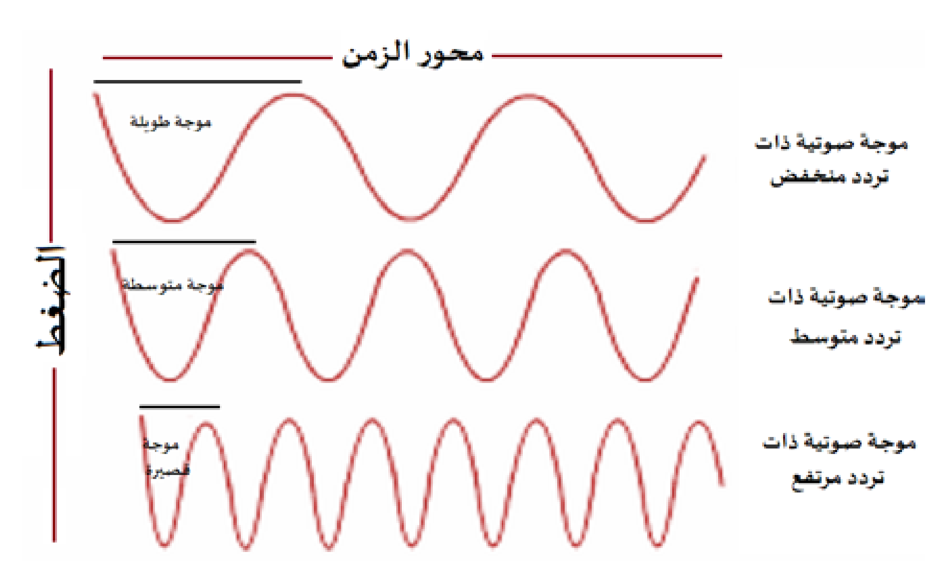


الشَّكل ١ : رسم تخطيطي للموجة الصوتية مشتمل على طول الموجة وسعتها والقيم الصفرية، موزعة على محوري الزمن والضغط

وصفة الدورية للموجة عبارة عن التردد أو التواتر، وهو مقدار تكرر الموجة الواحدة ذات الطول الموجي المتفق عليه في كل وحدة زمن.

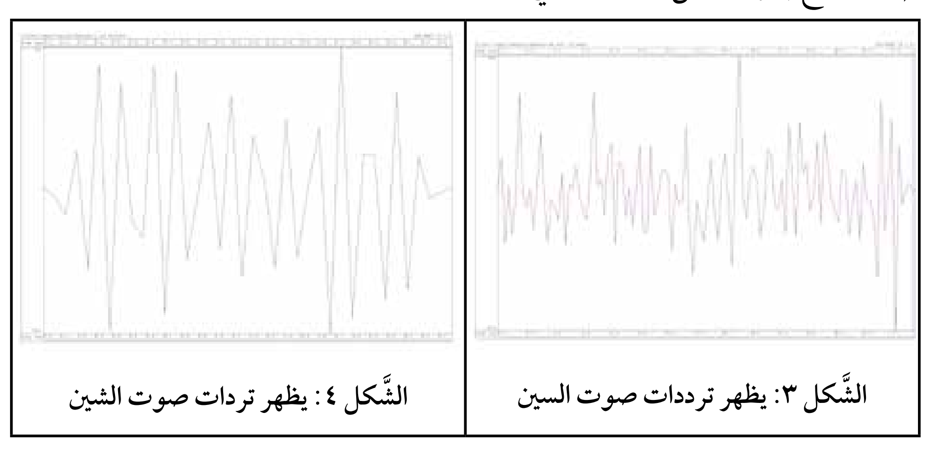
وتكون درجة الصوت أعلى كلما كانت الذبذبات أسرع، وكلما زادت سرعة هذه الذبذبات كان الصوت دقيقًا أو حادًّا، ومن ثم فإن الصوت يكون سميكًا إذا قل عدد الذبذبات في الثانية الواحدة. «فالفرق بين شوكة رنانة ذات درجة صوتية عالية وأخرى ذات درجة صوتية منخفضة أن الأولى تعمل عددًا أكبر من الذبذبات في الثانية الواحدة» (عمر، ١٩٩٧ : ٦٧). «ويقاس التردد بعدد الذبذبات التي تحدث في الثانية الواحدة» (الغامدي، ٢٠٠٠: ٣١).

وإذا كان تردد الصوت (Sound Frequency): هو عدد الذبذبات أو الموجات الكاملة في الثانية الواحدة (أيوب، ١٩٨٤: ٥٦)، فيمكننا رصد هذا التردد كما في (الشكل ٢) الذي يعرض لموجة صوتية ذات تردد مرتفع وأخرى ذات تردد منخفض.



الشَّكل ٢: موجة صوتية ذات تردد مرتفع (صوت رفيع) وموجة صوتية ذات تردد متوسط وموجة صوتية ذات تردد منخفض (صوت غليظ)

ونلاحظ هنا أن التردد يتناسب عكسيًّا مع طول الموجة، أي كلما زاد التردد قصر طول الموجة، أي: التردد ٤٠٠٠ هرتز له طول موجة أقصر من التردد ٥٠٠ هرتز. ويعتبر التردد من العناصر التي تؤدي إلى إدراك بعض الأصوات وتمييزها؛ حيث «يؤدي اختلاف تردد موجات بعض الأصوات اللغوية إلى تغير في إدراكها، فالصوت / س/ ذو تردد عال يفوق ٤٠٠٠ هرتز، فإذا انخفض تردده ليقترب من ٢٥٠٠ هرتز فإن السامع يدركه / ش/ » (الغامدي، ٢٠٠٠: ٧٨).



الشَّكل ٣: يظهر ترددات صوت السين

الشَّكل ٤ : يظهر تردات صوت الشين

٢,١,٣. سرعة الصوت (Speed of sound):

هي سرعة انتقال الطاقة الصوتية في الوسط، وهي ثابتة في الوسط الواحد بغض النظر عن نوع الصوت وتردده، ولكنها تختلف من وسط إلى آخر فتتغيَّر طبقًا لكثافة الوسط وإعاقته للصوت ودرجة الحرارة. وتتمثل خصائص الوسط التي تحدد سرعة الصوت في الكثافة وقابلية الانضغاط. وكلما زادت الكثافة وزادت قابلية الانضغاط، قلت سرعة الصوت. وتزداد سرعة الصوت في السوائل عن الغازات وفي الأجسام الصلبة عن السوائل؛ وذلك لتقارب الجزيئات بها، فالسوائل والأجسام الصلبة بصفة عامة أكثر كثافة من الهواء، وهي أيضًا أقل من الهواء بكثير في قابلية الانضغاط. ولذلك، فإن الصوت ينتقل بسرعة أكبر خلال السوائل والأجسام الصلبة.

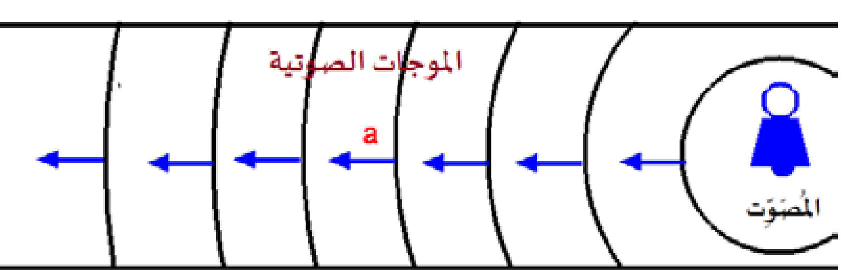
وعليه فإنَّ الصوت عبارة عن موجة طولية ينتجها ضغط المادة وتنافرها. وعمومًا فإنَّه ينبغي أنَّ نضع في الحسبان عند حساب سرعة الصوت مجموعة من الخصائص التي تؤثر على تلك السرعة مثل الضغط والكثافة والكتلة الجزيئية؛ فكلما قَلَّت كثافة الوسيط، زادت سرعة الصوت، وكلما زاد الضغط، كان الصوت أبطأ.

سرعة الصوت هي المسافة المقطوعة لكل وحدة زمنية بواسطة موجة صوتية أثناء انتشارها عبر وسيط مرن، عند ٢٠ درجة مئوية (ما يعادل ٦٨ درجة فهرنهايت) ، وتبلغ سرعة الصوت في الهواء حوالي ٣٤٣ مترًا في الثانية (١٢٣٤,٨ كم / ساعة) ؛ أو ١,١٢٥ قدمًا في ثانية ؛ أو ٧٦٧ ميلًا في الساعة. وتعتمد سرعة الصوت على درجة الحرارة، وعلى نسبة تشبع الهواء بالغازات. ويتم حساب سرعة الصوت في الهواء بواسطة الصيغة الآتية:

سرعة الصوت (م / ث) = ٣٣١,٥ + ٠,٦٠ ت (درجة مئوية)

وإذا كان الوسيط غير هوائي فإن تلك السرعة تكون أكثر تباينًا؛ حيث ينتقل الصوت بصورة أبطأ في الغازات؛ وبصورة أسرع في السوائل وفي المواد الصلبة. وعلى سبيل المثال إذا كانت سرعة الصوت في الهواء ٣٤٣ مترًا في الثانية فإنها تنضاعف في السوائل لتصل إلى ١٤٨٠ مترًا في الثانية، وتبلغ ٥١٢٠ مترًا في الثانية عبر الوسط الصلب مثل الحديد، بل وتبلغ في المواد شديدة الصلابة مثل الماس حوالي ١٢٠٠٠ مترًا في الثانية، أي حوالي ٣٥ ضعفًا من سرعته في الهواء (Bannon ،٢٠١٥).

سرعة الصوت



Speed of sound (a) depends on the type of medium and the temperature of the medium.

For calorically perfect air: γ_{perf} = ratio of specific heats = constant = 1.4

$$a^2 = R\ T\ \gamma_{perf}$$

For calorically imperfect air: γ = ratio of specific heats = γ (T)

$$a^2 = R\ T \left\{ 1 + \frac{(\gamma_{perf} - 1)}{1 + (\gamma_{perf} - 1)\left[\left(\frac{\Theta}{T}\right)^2 \frac{e^{\Theta/T}}{(e^{\Theta/T} - 1)^2} \right]} \right\}$$

Θ = 3056 °K

R = gas constant (286 $m^2/s^2/K°$ for air) T = absolute temperature °K

الشَّكل ٥: سرعة الصوت، المصدر (www.grc.nasa.gov).

٣, ١, ٣. شدة الصوت (Sound intensity):

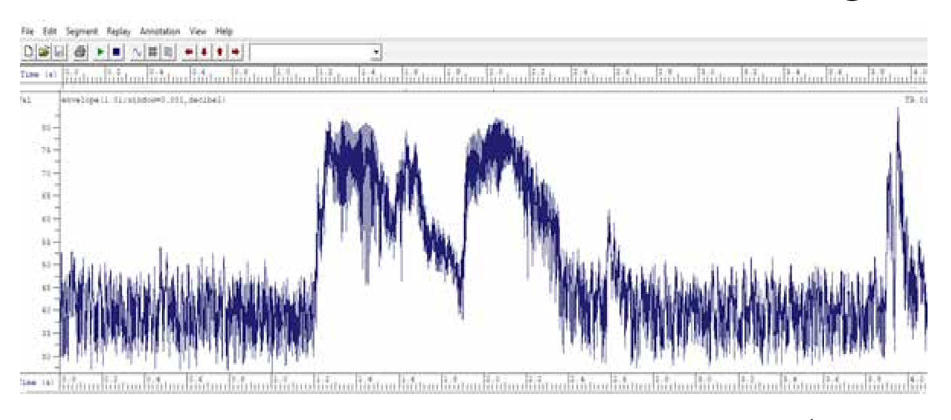
يُعَرَّفُ مستوى شدة الصوت والذي يشار إليه بالمصطلح (Sound intensity) أو (Acoustic intensity) بأنه القدرة أو الطاقة، التي تحملها الموجات الصوتية لكل وحدة مساحية في اتجاه عمودي لتلك الوحدة، والسياق المعتاد لقياس شدة الصوت هو الهواء. ويتم قياس شدة الصوت بوحدة الواط لكل متر مربع أو لكل سنتيمتر مربع (واط/ م ٢ أو واط / سم ٢)، وغالبًا يتم قياس شدة الصوت وفقًا لمعيار كثافة السمع القياسي الذي يشار إليه بالديسيبل (decibel) الذي تعبر عنه هذه المعادلة:

$$I_0 = 10^{-12}\ watts / m^2 = 10^{-16}\ watts / cm^2$$

والديسيبل هو وحدة شدة الصوت، وهو وحدة ليس لها تمييز؛ لأنها نسبة بين كميتين من الطاقة، وهو الأساس لقياس كل الأصوات في الطبيعة مثل الكلام، والضوضاء، وأصوات الحيوانات وغيرها، ويمثل الديسبل «مقياسًا عمليًّا لقوة الصوت وضغطه،

والديسيبل ١/ ١٠ بيل، والبيل هو ضغط أضعف صوت تدركه الأذن الإنسانية» (أيوب، ١٩٨٤: ١١٦).

ومن ثمَّ فإن مصطلح شدة الصوت (Sound intensity) لا يرادف مصطلح ضغط الصوت (Sound Pressure)، وإن كان هناك ترابط بينهما مرجعه أن الصوت المسموع يتكون من موجات ضغط، ومعلوم أنَّ إحدى الطرق المتبعة في تقدير شدة الصوت هي تحديد مقدار تباين الضغط الصوتي مقارنة بالضغط الجوي. ويعتبر الداين هو وحدة قياس الضغط. ومقداره ١/ ١٠٠٠ من الجرام، «وضغطة موجة ما هو عبارة عن عدد الداينات التي تضغط على مساحة سنتيمتر مربع واحد (أيوب، ١٩٨٤: ١١٧). ويوضح الشكل الآتي شدة الصوت.



الشَّكل ٦: منظور توضيح لشدة الأصوات المكونة لكمة (صوت) بنطق المؤلف

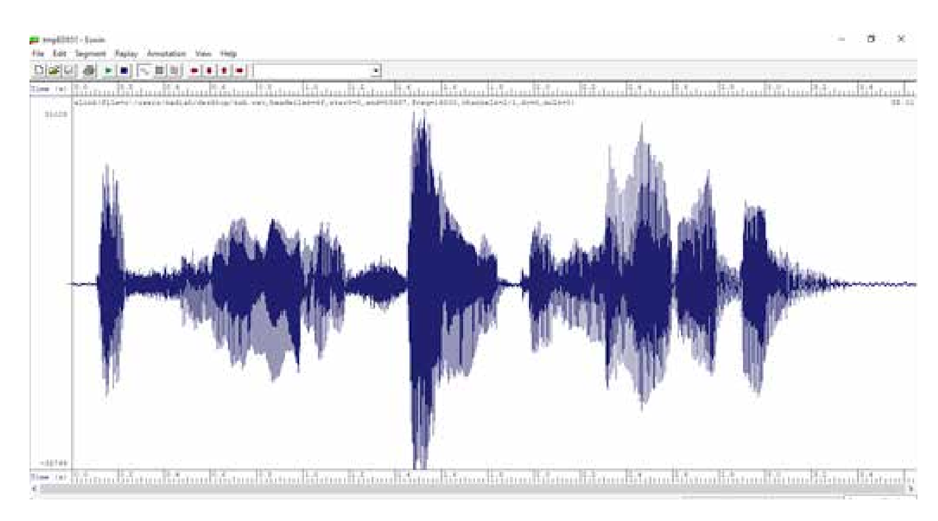
٤, ١, ٣. الرَّسم التَّذبذبيُّ (شكل الموجة) (Waveform):

الصوت عبارة عن: «اضْطِّرَاب مَاديّ فِي الهواء يَتَمَثَّلُ فِي قُوَّةٍ أَوْ ضَعْفٍ سَرِيعَيْنِ للضَّغْطِ المَتَحَرِّكِ مِنَ المصْدَرِ في اتجاه الخارج، ثم في ضَعْفٍ تَدْرِيجِيٍّ إلى نُقْطَةِ الزَّوَالِ النِّهَائِيّ». (الموسوي، ١٩٩٨: ٩٨) و (هلال، ١٩٨٨: ١٢٤). وعليه فإنَّ التوصيف الفيزيائي للصوت يتمثل في كونه سلسلة لتغييرات الضّغط في الوسط بين المرسل والمستقبل. فعندما يبدأ المرسل عملية إصدار الصوت عن طريق إنتاج مجموعة من الذبذبات التي تضغط على الكتلة الهوائية في شكل موجات متلاحقة تصطدم بالـمُسْتَقْبِل - ميكرفون أو أذن مثلًا - وهنا تحدث عملية تخلخل مصاحبة لهذا الضغط،

وأكثر الأمور شيوعًا لتمثيل معالم الصوت الحسية ورصدها هو الرّسم التّذبذبيّ، الذي يطلق عليه الشكل الموجي. ويمثل المحور الأفقي فيه محور الزّمن وتتم قراءته من اليسار إلى اليمين بينما يمثل المنحنى الرأسي قوة هذا الصوت في لحظة معينة، في الوقت الذي تظهر فيه عدد الترددات الذبذبية للحظة الواحدة على هيئة تتابع وتلاحق هذه الخطوط الرأسية لهذا المنحنى.

ويمكن الإشارة إلى الشكل الموجي (Waveform) على أنه مقابل منظور للصوت المنطوق، وعليه فإنَّه - أي الشكل الموجي - أقرب ما يكون إلى رسم بياني لمحور الجهد موزعًا على محور الزمن. وقد سبقت الإشارة إلى أنَّ الصوت ينشأ عن أشكال موجية تتعاقب بصورة دورية. وبالتالي فإنَّ الأدوات الإلكترونية والحاسوبية يمكنها رسم هذه الأشكال الموجية الدورية وعرضها على هيئة رسم بياني، يظهرها على شكل أنواع من الخطوط المتعرجة. وتتم قراءة هذا الشكل الموجي من اليسار إلى اليمين عبر محور الزمن الأفقي، في الوقت الذي يعرض فيه المحورُ الرأسيُّ - من أعلى إلى أسفل - إشارات السعة الصوتية أو ما يمكن تسميته بالجهد النسبي (Relative Voltage).

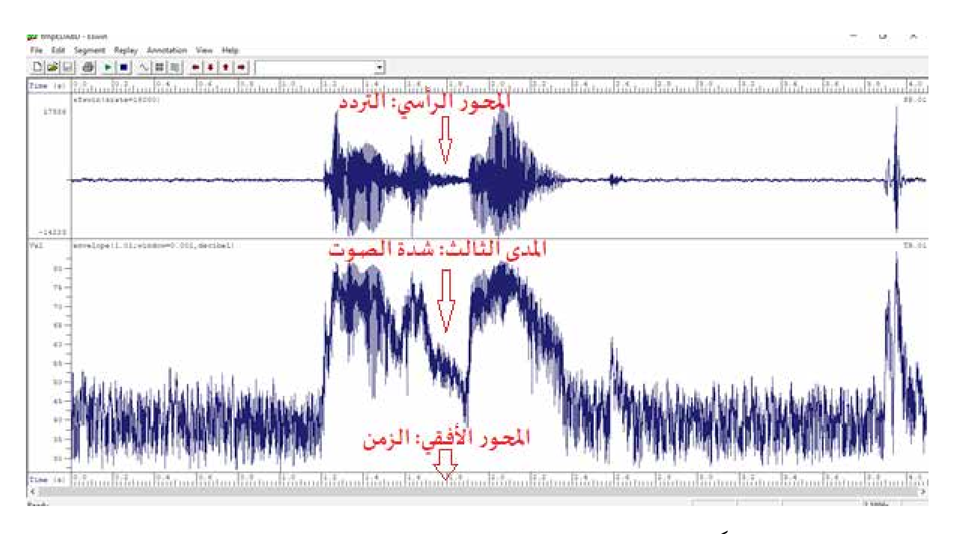
فعندما ننطق أية كلمة (دفقة صوتية واحدة) في البداية سيتم التنفس من الرئتين ثم يخرج الهواء إلى الحنجرة الذي قد يسبب عند مروره بها اهتزاز الحبلين الصوتيين ويستمر في مسيره حتى يصل إلى القناة الصوتية وينتقل منها إلى التجويف الفموي، وفي الأصوات الأنفية يضطر الهواء للمرور بالتجويف الأنفي (nasal cavity) وبعدها تخرج الكلمة وتصبح مسموعة وتتحول إلى شكل الموجة. ويظهر (الشَّكل ٧) صورة للرسم التذبذبي من خلال برنامج (SFS).



الشَّكل ٧: الرَّسم التَّذبذبيُّ لجملة (دراسة الصوت اللغوي) كما يعرضه برنامج (SFS)

٥, ١, ٣. الصورة الطيفية (Spectrogram):

الصورة الطيفية للصوت عبارة عن تمثيل مرئي ثلاثي الأبعاد لتردد الإشارات الصوتية وتَغَيُّرِهَا عبر محور الزمن وشدتها، ويمثل المحور الأفقيُّ في الصُّورة الطَّيفيَّة منحنى الزَّمن بالمِيللي ثانية، ويمثل التردد المحور الرأسي بالهيرتز، أما البعد الثّالث وهو مدى التّردّد أو شدة الصوت فيُمَثِّله قرب أو بعد لون الصورة عن لون الخلفية. وقد تعتبر الصّورة الطّيفيّة عددًا من السّلاسل المتوالية إذا نظرنا إليها من أعلى (مسقط أفقي)؛ حيث تُمَثَّل الزّيادات أو القمم الرأسية في الصور الطيفية بالبقع المظلمة في الصّورة الطّيفيّة، وكلتاهما تعبر عن شدة الصوت المنطوق.



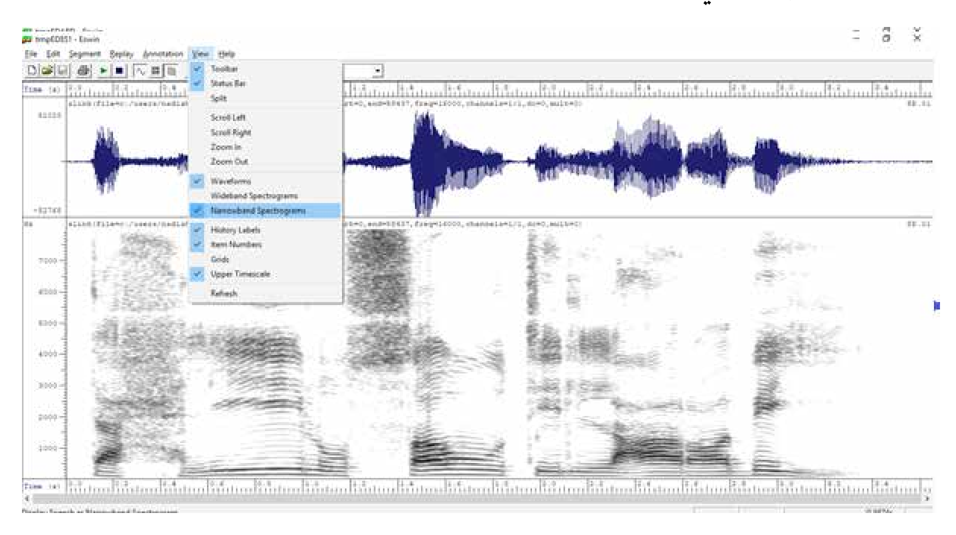
الشَّكل ٨: مكونات الصورة الطيفية: التردد والشدة والزمن

وأحيانا يطلق على الصورة الطيفية مسميات مماثلة أو مقاربة مثل مسمى المطبوعات الصوتية (voiceprints) أو المخطط الصوتي (voicegrams) أو الصورة الصوتية (sonographs) وهي عبارة عن أداة تحليلية للصوت ينتج عنها رسوم تُمَثِّلُ الصوت حسب الترددات المكونة له. وكان استخدام الصورة الطيفية للصوت محصورًا بادئ الأمر في مجالات الموسيقى، والسونار، والرادار، وقياس الزلازل، وما لبث أنَّ تطوَّر استخدامها في مجال معالجة الكلام (JL Flanagan, 1972: 23)، ثم ذاع استخدامه في مجال إنتاج الكلام وتوليدة وتحليله بصورة آلية.

وتسمى الآلة المنتجة للصور الطيفية المطياف البصري (Optical Spectrometer)، ويعتبر المطياف واحدًا من أهم الأجهزة التي ساعدت في رصد الخواص الأكوستية للصوت اللغوي وتحليله؛ حيث يقوم بعرض صورة مرئية للصوت المنطوق كاشفة عن خصائصه، وقد عوَّل عليه علماء الصوتيات في تحديد صفات الأصوات اللغوية وترجيح أوجه الخلاف في وصف بعض الأصوات التي دار حولها الجدل قديمًا وحديثًا، «ونظرًا لأن موجات الصوت اللغوية من النوع المركب فإن عرضها باستخدام جهاز عارض الذبذبات لا يقدم كل التفاصيل عن الموجة الصوتية. بينما يقدم المطياف ثلاثة أبعاد للموجة المرسومة وهي: التردد، والشدة، والزمن. وهذا يعين الباحث على معرفة زمن الصوت، والتردد الأساسي، والنطاق الرنيني وشدته» (الغامدي، ٢٠٠٠: ١١٤).

وقد ظلت معظم دراسات أكوستية الصوت اللغوية تعتمد على المطياف بشكل أو بآخر لمدة أربعة عقود حتى وجدت البديل الأمثل في البرامج والأدوات الحاسوبية.

فالبرامج الحاسوبية الآن يمكنها عرض تلك الصور الطيفية بطريقة أكثر دقة وسهولة وفاعلية، بل وتحويل بيانات تلك الصور إلى قِيَم رقمية ثابتة ومحددة، ومن أمثلة هذه البرامج (Cool edit pro) وبرنامج (Gold wife)، وبرنامج (Speech filing System) (البهنساوي، ٢٠٠٤: ١١).



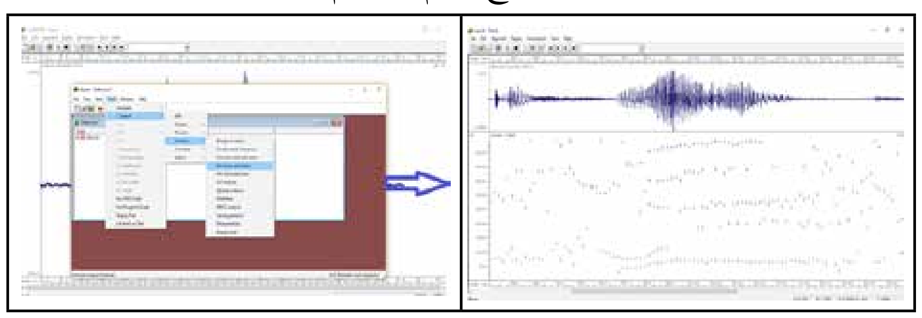
الشَّكل ٩: كيفية عرض الصورة الطيفية لجملة [تحليل الصوت العربي] من خلال برنامج (SFS)

ونستطيع أن نقرأ مجموعة من الخصائص الأكوستية للأصوات التي تم نطقها ومن ثَمَّ عرضها في هذه الصورة الطيفية، ويمكننا ملاحظة اختلاف طبيعة الأصوات المنطوقة بناءً على معطيات الصورة الطيفية، فالصوت الأول في الجملة هو صوت التاء، وهو صوت مهموس انفجاري، ويظهر أمامنا كغيره من الأصوات الانفجارية المهموسة التي تبدأ بفترة صامتة يحتبس فيها النَّفَس، وبعدها بحوالي ٨٢ ميللي ثانية تظهر طاقة الطّاقة فجأة وبقوة في نطاقات التّردّد أو الحزم، على هيئة انفجار. وبعدها تظهر الفتحة القصيرة مستغرقة زمنا مقداره ٩٨ ميللي ثانية بنطاقات تردد عالية ناتجة عن صفة الجهر التي تتسم بها كل الحركات طويلها وقصيرها بالإضافة إلى بعض الصوامت، ثم يبدأ حرف الحاء ابتداءً من الخط الزمني ٠,١٨٠ ثانية وحتى الخط الزمني ٠,٣٧٠ ثانية،

وهو صوت احتكاكيّ مهموس تكون الطّاقة عند النطق به مركّزة بكثرة في نطاق التّردّد أو الحزم، وفوضويّ إلى حدّ ما (شبيه بالضّوضاء) في مظهره.

أما الأصوات المجهورة مثل اللام والياء والواو والعين والغين مثل فتبدو أكثر تنظيمًا، وتبدو أعلى السّلسلة (البقع المظلمة) في الواقع على هيئة خطوط في وضع أفقيّ الشّكل عبر الصّورة الطّيفيّة. هذه الحزم الصوتية (المعالم formants) هي الترددات أو مجموعة الترددات (Groups of frequencies) التي تشكل نوع الصوت (Tamper) وتميزه عن الأصوات الأخرى ذات الأنواع المختلفة. وتمثّل هذه الحزم الذّبذبات (عمر، ١٩٩٧: ١١٦)؛ حيث يعطي شكل الفم رنين الصوت، ويتم ترقيمها من الأسفل إلى الأعلى على هذا النحو: م١ م٢ م٣ أو F1, F2, F3 إلخ؛ وتختلف مواضع هذه الحزم باختلاف نوع الصوت المنطوق، ومن الممكن تحديدها بشكل تقريبي مع كل فونيم.

ومن خلال (الشَّكل ١٠) نرى توزيع المعالم أو الحزم الصوتية لكلمة مفردة.



الشَّكل ١٠: يوضح طريقة استخراج الحزم الصوتية من خلال برنامج (SFS)؛ بينما يوضح الشَّكل ١١: أبعاد الحزم الصوتية لكلمة [صوتيات]، التي سُجِّلَت بصوت المؤلف

٢,٣. التباين الأدائي:

تتباين مستويات الكلام البشري بين مختلف اللغات واللهجات التي يمارسها المجتمع الإنساني تعبيرًا عن أفكارهم وأغراضهم، وتتصف اللغة بشكل عام بخاصيتها الديناميكية التي تؤدي إلى تنامي مكوناتها واتساعها بشكل مضطرد لاسيما على مستوى المفردات التي تعد واحدة من أهم لبنات اللغة، وإذا انتقلنا إلى المستوى الصوتي الذي يعد محور حديثنا فسنجد أن ذلك التباين يصبح أشدَّ تعقيدا وأكثرَ فردية، فلكل فرد في المجتمع سمات صوتية فسيولوجية وأدائية تميزه عن غيره من ناطقي اللغة، وهذا ما

يجعل البحث في قضايا التحليل الصوتي الحاسوبي أكثر اتساعًا عن غيره من مستويات التحليل اللغوي، لاعتماده على التباين اللغوي الواضح بين الناطقين باللغة، حيث تعتبر دراسة أثر اختلاف الجنس واللكنة والعمر من القضايا المهمة في التعرف على الكلام المنطوق (Huang,2001: 8).

وتتحدد ملامح هذا التباين الصوتي في المدى الزمني (Duration) والنغمة الصوتية (Pitch) والتردد الأساسي (Fundamental Frequency) والأداء الشخصي أو الذاتي الثابت أو المتغير لدى المتحدث أثناء عملية النطق (Speaker Variability).

١ , ٢ , ٣. المدى الزمني (Duration):

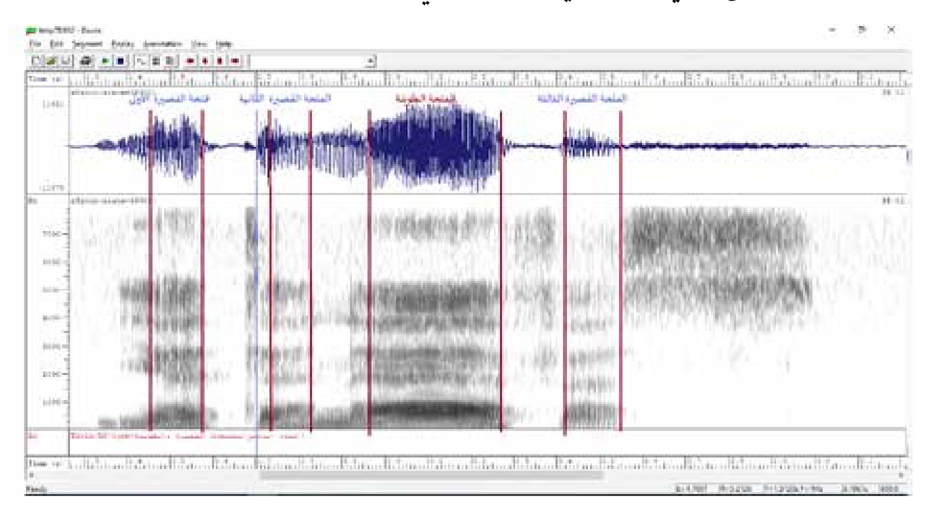
يمثل الطول أو المدى الزمني للفونيم أحد أهم السمات التي تلعب دورًا جوهريًّا في التفرقة بين صوت وآخر، ويختلف هذا المدى الزمنى طولًا وقصرًا وفقًا للسمات الصوتية (Khattab, 2014: 235)، فهناك فارق ملحوظ بين زمن النطق بالأصوات المهموسة عن زمن النطق بالأصوات المجهورة، وكذلك يظهر فارق واضح بين زمن النطق بالأصوات الرخوة عن زمن نطق الأصوات الشديدة، فلكل مجموعة صوتية حيز متقارب من المدى الزمني، وتعد الحركات الطويلة والصوامت المضعفة أو (المشددة) والتي يطلق عليها (geminates) من أطول الأصوات في اللغة العربية (Blust, 2013: 17). ومعلوم أن الحرف المضعَّف أو المشدد في اللغة العربية عبارة عن مزج بين حرفين متماثلين أولهما ساكن والآخر متحرك، ويرمز لهذا التضعيف برأس سين توضع أعلى الحرف المشدد، ومثالها حرف الدال في كلمة (شَدَّ). وظاهرة التشديد ليست ظاهرة مقصورة على اللغة العربية، بل هي سمة لغوية في لغات عدة مثل: اللغة العبرية والبربرية والمالطية والدنماركية والإستونية والفنلندية والهندية والهنغارية والإيطالية واليابانية واللاتينية والتاميلية. وهاتان الظاهرتان (الحركات الطوبلة والتشديد) مترابطتان في بعض اللغات مثل: الإيطالية والإستونية والنرويجية؛ إلا أنهما غير مترابطتين في لغات مثل العربية واليابانبة والسويدية.

وقد أبدى علماء العربية اهتمامًا بالغًا بالحركات طويلها وقصيرها. ذلك أن لها تأثيرًا على الدلالة، فقسموها باعتبار زمنها إلى حركات قصيرة وهي الفتحة والضمة والكسرة، وحركات طويلة طويلة وهي ما أطلفوا عليه (حروف المد) بأنواعها الثلاثة:

الأَلِف السَّاكِنَة الـمَفْتُوحُ مَا قَبْلَهَا دائمًا (ـَا)، وَالوَاوُ السَّاكِنَةُ الـمَضْمُومُ ما قَبْلَهَا (ـُو)، وَالياءُ السَّاكِنَةُ الـمَكْسُورُ ما قَبْلَهَا (ـِي) (أحمد، ٢٠٠٤: ٢٨٦)، كما تضم اللغة العربية مجموعة من أنصاف العلل متمثلة في الياء والواو المفتوح ما قبلهما، بالإضافة إلى أصوات العلة وهي الياء والواو المتحركتان.

وقد ردَّ علماء التجويد العرب سبب الاختلاف بين الحركة الطويلة من ناحية وبين الحركة القصيرة والصوامت المشددة من ناحية أخرى، مستندين إلى عاملين أحدهما اتساع المدى الزمني والآخر الامتداد النطقي، «فإذا استوفى حرف المد نصيبه من المد انتقل بذلك من الحركة إلى الحرف، وهذه الخاصية ثابتة لحروف المد دون غيرها من الأصوات الجامدة «لا سيما الشديدة (أي الانفجارية) فإنها آنية الحدوث، وكذلك الرخوة (الاحتكاكية) فإنها وإن كانت زمانية يمتد بها الصوت مدة، لكن ذلك الامتداد لا يبلغ مقدار ألف، أي مقدار نطق حرف المد» (الحمد، ٢٠٠٧: ٥٣٦).

ويُظهر الشكل التالي الفارق في المدى الزمني بين الفتحة القصيرة والفتحة الطويلة.

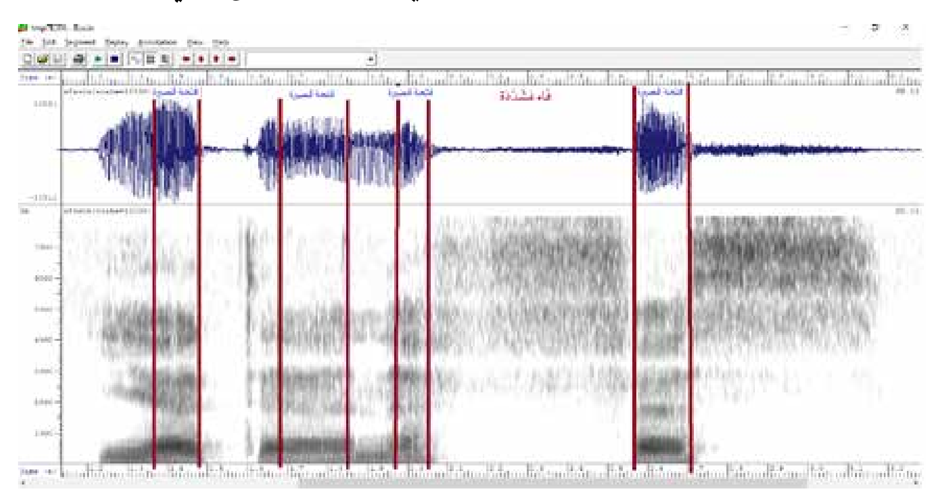


الشَّكل ١٢: الفارق في المدى الزمنى بين الفتحات القصيرة والفتحة الطويلة (ألف المد) في كلمة (يتنافس) بصوت المؤلف

وبالنظر إلى (الشكل ١٢) نجد أن الفتحة الأولى التي أعقبت صوت الياء بدأت من الثانية ١,٤٧١ وحتى الثانية ١,٥٧٢ مستغرقة زمنًا قدره ٠,١٠١ ميللي ثانية. في حين أنَّ الفتحة الثانية التي أعقبت صوت التاء بدأت من الثانية ١,٧٠٧ وحتى الثانية

١,٨٠٩ مستغرقة زمنًا قدره ١٠١,٠ ميللي ثانية. بينما جاءت الفتحة الطويلة المتمثلة في ألف المد مستغرقة زمنًا قدره ٣٥٢,٠ ميللي ثانية، حيث بدأت من الثانية ١,٩٥٢ وحتى الثانية ٢,٢٧٨. وأخيرًا جاءت الفتحة الثالثة التي أعقبت صوت الفاء مبتدئة من الثانية ٢,٤١٣ وحتى الثانية ٢,٥٣٧ مستغرقة زمنًا قدره ١٢٣,٠ ميللي ثانية.

وإذا ما أردنا أن نقارن بين زمن الحركات من ناحية وبين المدى الزمني للحرف المشدد فإننا سنجد فارقًا ملحوظًا على النحو الذي يوضحه الشكل الآتي:

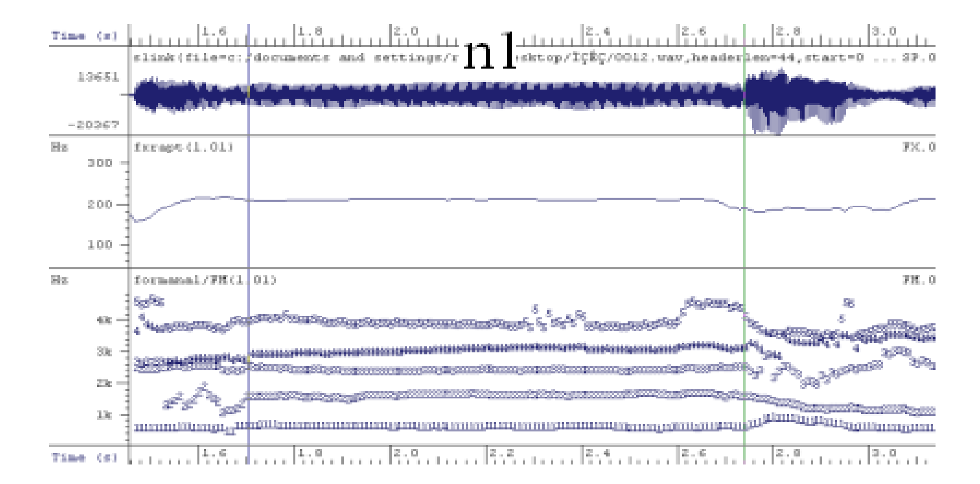


الشَّكل ١٣: الفارق في المدى الزمني بين الفتحات القصيرة والحرف المشدد في كلمة (يَتَنَفَّسْ) بصوت المؤلف

وبالنظر إلى (الشكل ١٣) نجد أن الفتحة الأولى التي أعقبت صوت الياء بدأت من الثانية ١,٣١٤ وحتى الثانية ١,٤٧٧ مستغرقة زمنًا قدره ١٦٣,٠ ميللي ثانية. في حين أنَّ الفتحة الثانية التي أعقبت صوت التاء بدأت من الثانية ١,٦٩٢ وحتى الثانية ١,٨٤٤ مستغرقة زمنًا قدره ١٥١,٠ ميللي ثانية. ثم بدأت الفتحة الثالثة التي أعقبت حرف النون في الثانية ١,٩٤٧ وانتهت عند الخط الزمني ٢,٠٦٨ مستغرفة زمنًا قدره ١٢٠,٠. بينما جاء صوت الفاء المشدد مستغرقة زمنًا قدره ٥٢١,٠ ميللي ثانية أي بزيادة ١٦٩,٠ وهو ما يمثل نسبة مئوية تقارب ١٤٨ ٪، حيث بدأت من الثانية ٢,٠٤٢ وحتى الثانية ٢,٥٦٣. وأخيرًا جاءت الفتحة الثالثة التي أعقبت صوت الفاء مبتدئة من الثانية ٢,٥٦٣ وحتى الثانية ٢,٦٩٢ مستغرقة زمنًا قدره ١٣١,٠ ميللي ثانية.

وقد لفت نظري أثناء تحليلي لِلْأَمْدَاءِ النسبية للأصوات العربية اتساع المدى الزمني مع الصوامت المشددة بشكل عام، ومع صوتي النون والميم المشددتين في الأداء القرآني بشكل خاص، وللتأكد من ذلك قمت بجمع قاعدة بيانات احتوت على ٢١٥ جملة صوتية متنوعة من القرآن الكريم، اشتملت على سورة الفاتحة، وما تيسر من سورة البقرة والكهف وبعض سور جزء عم، وقد اشتملت قاعدة البيانات هذه على ١٠٥٢٧ فونيمًا، ويمكننا حصر نتائجها فيما يتعلق بالنون والميم المشددتين على هذا النحو:

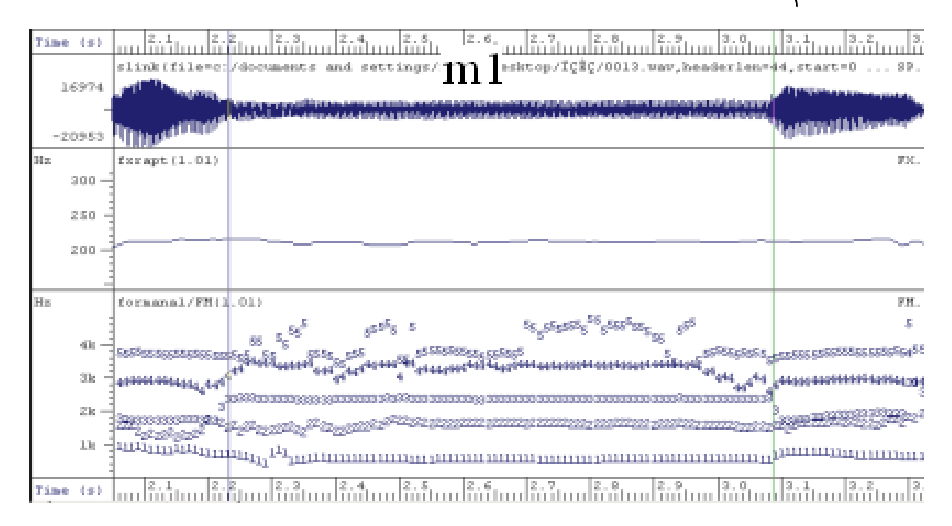
أولًا: النون المشددة



الشَّكل ١٤: صورة طيفية للنون المشددة، وتظهر فيها ثلاثة مستويات من مستويات التحليل الطيفي، يبدأ من الـ (*wave form*) ويتوسطه منحنى التنغيم (*Fundamental Frequency*)، وينتهي بمنحنى الفورمانت (*f1,f1,f3..*)



ثانيًا: الميم المشددة:



الشَّكل ١٥: صورة طيفية للميم المشددة، وتظهر فيها ثلاثة مستويات من مستويات التحليل الطيفي، يبدأ من الـ (*wave form*) ويتوسطه منحنى التنغيم (*Fundamental Frequency*)، وينتهي بمنحنى الفورمانت (*f1,f1,f3..*)

وإذا ما تأملنا تحليلنا لصوتي الميم والنون المشددتين فسنلمح أنَّ كليهما صوتٌ احتكاكيٌّ بينيٌّ مجهورٌ أَغَنٌّ مُرَقَّقٌ، وقد وردت النون المشددة في هذه النماذج ٦٩ مرة، وبلغ متوسط مداها الزمني حوالي ٥٤٢ ميللي ثانية، وهذا المدى الزمني أكثر من ثلاثة أضعاف زمن النون الساكنة، حيث يبلغ متوسط مداها الزمني الطبيعي ١٥٩ ميللي ثانية.

أما الميم المشددة فقد وردت في هذه النماذج ٥٩ مرة، وبلغ متوسط مداها الزمني حوالي ٥١٠ ميللي ثانية، وهذا المدى الزمني يقترب من ثلاثة أضعاف زمن الميم الساكنة، حيث يبلغ متوسط مداها الزمن الزمني ١٦٩ ميللي ثانية.

وتلعب معرفة المدى الزمني للفونيم دورًا مهمًا في فهم الكلام، إلا أن معظم أدوات التعرف على الصوت آليًّا تتجاهل هذا الدور، بسبب الاعتماد على أكثر النماذج شهرة في التعرف على الصوت المنطوق/ وهي نماذج (Hidden Markov Models) والتي تعرف اختصارًا بـ(HMMs). وهي نماذج إحصائيَّة مخفيَّة تعبر عن أبسط شبكة ديناميكية. وطورها كل من L. E. Baum وزميله Coworkers. وتعتمد هذه النماذج على الجانب الكيفي في توصيف الفونيم أكثر من اعتمادها على الجانب الكمي (L. E. Baum, 2017: 27).

٢,٢,٣. النغمة الصوتية (Pitch):

التنغيم ظاهرة صوتية عامة، وكثيرًا ما تؤدي هذه الظاهرة إلى توجيه دلالة الخطاب دون الحاجة إلى تغيير مفردات الجملة، «ويؤدي التنغيمُ دورًا فاعلًا في التقريرِ والتوكيدِ والتعجبِ والاستفهامِ والنفيِ والإنكارِ والتهكمِ والزجرِ، وغيرها من أنواع الفعل الإنسانيّ كالغضب واليأس والفرح والحزن بوساطة التلوين في الدرجات التنغيمية بمستوياتها العليا والمتوسطة والهابطة، ولذلك عدّها علماءُ اللغةِ من الفونيمات غير التركيبية التي من شأنِها أن تعرفَنا على مواقف المتكلمين من خلال تنوع ظهورها» (محمد، ٢٠٠٩: ٩).

وقد دأب علماء الأصوات على تناول التنغيم في إطار وصفه بأنه فونيم غير تركيبي، ويقصدون: الظواهر الصوتيةٌ التي لا تتشكل ضمن حيزٍ زمني مستقلٍ، «فالتنغيم ليس «جزءًا من التركيبِ اللغويِّ في الجملة، بل هو حدثٌ طارئٌ على التركيبِ يصاحبه، ويتغير نتيجة تغيره في السياق اللغوي الجاري فيه، إذ يربطُ التنغيمُ عناصرَ التركيبِ بعضها ببعض» (عطية، ٢٠١٥: ٢٣).

فالتنغيم الكلامي بمثابة مقامات موسيقية يتبعها المتحدث لإضافة لمسة فنية على خطابه، فحين ينطق يتكلم الإنسانَ «لا يتبعُ درجةً صوتيةً واحدةً في النطقِ بجميع الأصوات، فالأصواتُ التي يتكوّنُ منها المقطعُ الواحدُ قد تختلفُ في درجةِ الصوتِ وكذلك الكلماتُ قد تختلفُ فيها، ومن اللغاتِ ما يجعلُ لاختلافِ درجةِ الصوتِ أهميةً كبرى، إذ تختلفُ فيها معاني الكلمات تبعًا لاختلاف درجةِ الصوتِ حين النطق بها، ويمكن أن نسمي نظام توالي درجات الصوت بالنغمة الموسيقية» (أنيس، ١٩٦١: ١٧٥). وقد دأب الباحثون على تكسير مستويات التنغيم إلى ثلاثة أضرب:

١. النغمة الصاعدة: وتعني وجودَ درجةٍ منخفضةٍ في مقطعٍ أو أكثر تليها درجة أكثر علوًّا منها.

٢. النغمة الهابطة: وتعني وجودَ درجةٍ عاليةٍ في مقطعٍ أو أكثر تليها درجة أكثر انخفاضا.

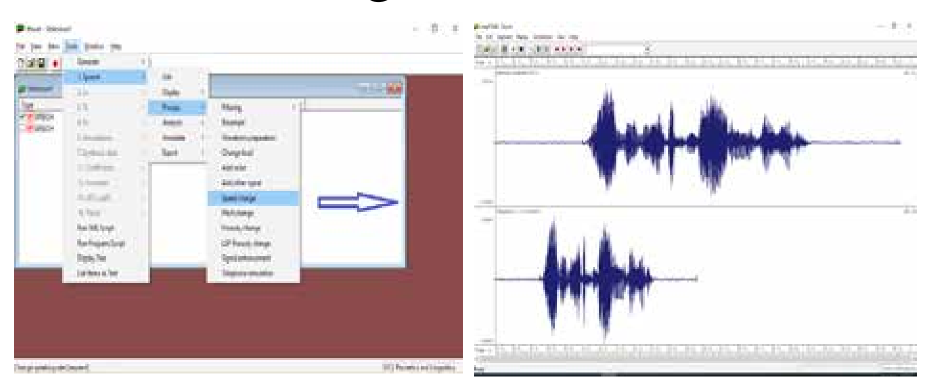
٣. النغمة المستوية: وتعني وجودَ عددٍ من المقاطعِ تكون درجاتُها متحدة.

إنَّ الصوت البشري أداة سحرية. فعن طريقه يمكننا تمييز من نعرفهم؛ وتسمح لنا الإمكانات الصوتية بتخليق نغمات غنائية صوتية تؤسر الأسماع؛ كما أن تنوع النغمات الصوتية تمكننا من التعرف على المشاعر الإنسانية المصاحبة للتواصل اللفظي؛ بالإضافة أنَّ تلك النغمات تمثل السمة الأولى لتكوين البصمة الصوتية التي تميز كل فرد عن بقية الأفراد الناطقة بنفس اللغة. حيث يتم إنشاء الصوت البشري الصوت من خلال عدد لا يحصى من حركات العضلات، الأمر الذي يترتب عليه أن يتألف هذا الصوت من العديد من المكونات المختلفة والمتفردة (different components)، وهي النغمة (pitch) و التنغيم (tone) والسرعة (rate).

وتمثل النغمة الصوتية أحد أهم مكونات الصوت الإنساني، فهي جزء لا يتجزأ من الصوت، ويتم تعريفها على أنها معدل اهتزازات التدفقات الصوتية، وكلما زاد معدل هذه الذبذبات كلما ارتفعت درجة الصوت، وبالتالي ينشأ عنه صوت حاد، وبالتالي فعندما تكون معدلات هذا الاهتزاز بطيئة فإنه ينشأ عنها أصوات أكثر عمقًا، وهو أمر مرتبط بشكل وثيق بمدى سُمْكِ وطُولِ الرقيقتين الصوتيتين (vocal cords) من ناحية، وبمدى شد أو جذب العضلات المحيطة بهما من ناحية أخرى (Stevens, ٢٠٠٠: ١٨). وتلعب الرقيقتان الصوتيان هنا الدور الأهم في تشكيل الذبذبات الصوتية، وتُعْرَفُ الرقيقتان الصوتيتان أيضاً بالحبال الصوتية، وهي عبارة عن فتيلتين متماثلتين من الأنسجة المكونة من الغشاء المخاطي، وتمتد أفقيًّا من الخلف إلى الأمام عبر الحنجرة، وتقوم بعملية تعديل تدفق الهواء القادم من الرئتين، في الوقت الذي تهتز فيه هاتان الرقيقتان أثناء مرور هذا الهواء المنبعث من الرئة، والتي تلعب دورًا أساسيًّا في عملية تكوين الأصوات أثناء النطق. ويؤثر حجم الحبال الصوتية على طبيعة الصوت.

وهذا ما يفسر سبب ارتفاع أصوات النساء عمومًا عن أصوات الرجال؛ ذلك نظرًا لطبيعة تكوين الرقيقتين الصوتيتين لديهن والتي تتميز بالقِصَر مقارنة بمثيلاتها عن الرجال، ومع ذلك فليست هذه هي العوامل فقط هي التي تؤثر على النغمة الصوتية؛ حيث تتأثر نغمة الصوت بالعواطف والحالات المزاجية والنفسية. فعندما تعتري الإنسان حالة خوف أو إثارة تتقلص العضلات حول الحنجرة (larynx) بشكل لا شعوري أو لا إرادي، ويترتب على ذلك مزيدٌ من الضغط على الرقيقتين الصوتيتين، فترتفع النغمة

الصوتية. والأمر لا يقتصر على العامل اللا إرادي؛ بل قد يحدث ذلك بطريقة إرادية في حالة التمرين الصوت والإلقاء وتقليد الأصوات، أو محاكاة بعض الأصوات البشرية أو غير البشرية. والأمر لم يعد يقتصر على العاملين الفسيولوجي والنفسي فقط، بل إن الأدوات الحاسوبية أتاحت لنا تغيير درجات النغم وسرعته، وأصبح بالإمكان تحويل أصوات ذكورية إلى أصوات أنثوية عن طريق البرامج الحاسوبية.



الشَّكل ١٦: طريقة إدخال تعدبلات على سرعة النغمة الصوتية من خلال برنامج (SFS)

٣,٢,٣. التردد الأساسي (Fundamental Frequency):

غالبًا ما يتم تعريف التردد الأساسي بشكل عام على أنه أدنى تردد في الشكل الموجي (waveform)، وفي الدراسات الصوتية يتم تعريفه بأنه مقياس مدى ارتفاع أو انخفاض تردد صوت الفرد، وغالبًا ما يرتبط هذا الارتفاع أو الانخفاض بالنغم (pitch) الذي يمثل توتر الحبلين الصوتيتين (vocal fold vibration) نتيجة اندفاع الهواء من الرئتين مما يؤدي إلى اهتزازات صوتية. وتختلف قيم هذا التردد الأساسي بين الذكور والإناث، فعلى سبيل المثال تبلغ قيمه عند الذكور البالغين ما بين ٨٥ و١٨٠ هيرتز في حين تبلغ قيمه عند الإناث البالغات ما بين ١٦٥ و ٢٥٥ هيرتز . هذا بشكل عام أثناء النطق العادي في الظروف العادية، أما عندما يصاحب هذا النطق تنغيما خاصًّا أو تجويدًا أو تغنيًا فإن هذا النطاق سيكون أكثر اتساعًا بنسبة تتوافق مع درجة النغم. ولا يقتصر التردد الأساسي في التمييز بين أصوات الذكور والإناث بل يمثل في الوقت نفسه أحد العوامل الفردية التي تميز صوت متحدث عن آخر، والذي يمكن عن طريقه إنتاج أصوات يختلف الواحد منها عن الآخر. وحين تسمع أي صوتين يمكن أن تقارنهما من

هذه الجوانب المختلفة، مثل شوكة رنانة وأرغن، فهما يصدران صوتين مختلفين نتيجة عامل أو أكثر» (عمر، ١٩٩٧: ١٧٩).

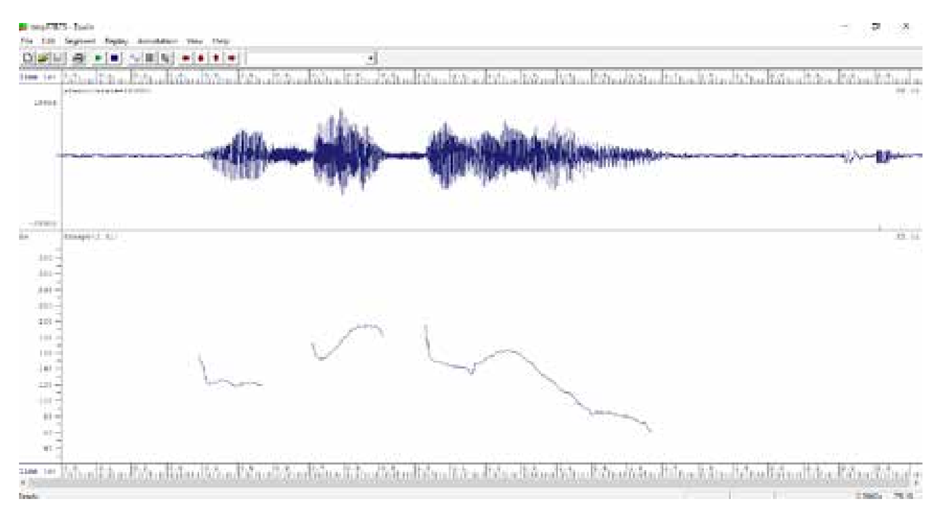
ويعبر شكل الموجة عن قيم شكل الموجة أو التردد الأساسي لهذا الصوت، فالصوت ينظر إليه عادةً على أنه عمليّة فيزيائية مادّيّة تتكوّن من جزأين أساسيين:

- منتج حقيقي للصوت؛ ويمثله الحبلان الصوتيان.
- منتج أو مشكل فرعي للصوت؛ وهو عبارة عن منتج مساعد، يقتصر دوره على إتمام عملية التوجيه واكتساب الصفات التمييزية، وتمثله باقي أعضاء النطق مثل: اللّسان والشّفاه والأسنان... إلخ.

وتحاول عملية تحليل التنغيم رصد هذا التردد الأساسي للصوت المنطوق. وهناك عدّة قواعد لمعرفة التردد الأساسي تخلص إلى أنه العنصر الأقوى الذي يجعلنا ندرك علاقة التلازم بين الصوت وصاحبه، فبمجرد أن نسمع جملة نعرف قائلها بسبب معرفتنا لتنغيمه. «ويمكن أن نميز صوتًا عن صوت آخر بشكل الموجة التي تُنْتِجُ كلًّا منهما، وليس شكل الموجة مقياسًا بالمعنى الدقيق، ولكنه مجرد وسيلة لتمييز الأصوات بتحليلها إلى الموجات التي تتكون منها» (أيوب، ١٩٨٤: ١٨٠).

وتظهر معالم التردد الأساسي على هيئة منحنيات أفقية موازية لخط الزمن، وتكون أكثر وضوحًا مع الأصوات المجهورة؛ حيث يهتزُّ الحبلان الصّوتيّان، وتتواجد القيم لهذه الحزم بين ٨٠-٢٠٠ هرتز بالنسبة للمتحدثين الذكور، وتتزايد هذه القيمة مع الإناث لضيق الحنجرة عن حنجرة الذكور، وقلة سمك الأحبال الصوتية؛ فتبلغ قيمة هذه الحزم مع الأنثى البالغة ١٥٠-٣٥٠ هرتز.

ويوضح (الشَّكل ١٧) منحنى التردد الأساسي لكلمة مفردة.

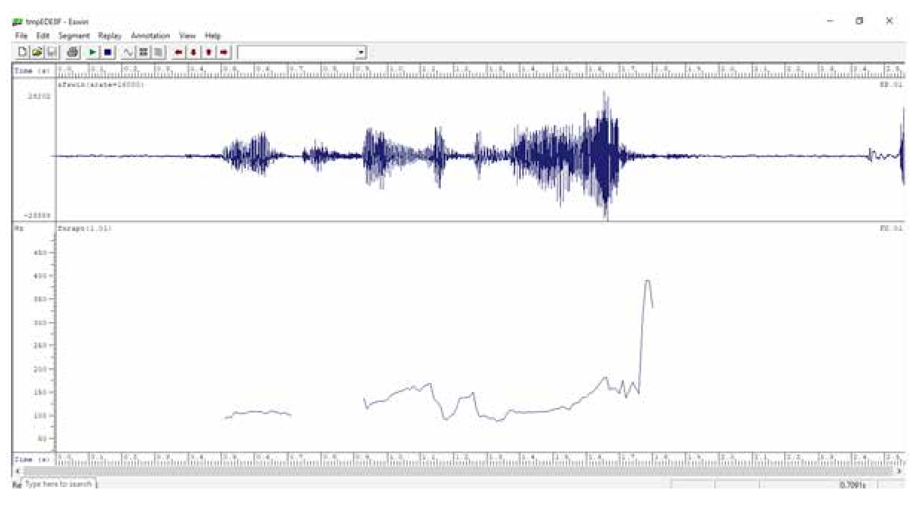


الشَّكل ١٧: منحنى التردد الأساسي لكلمة مفردة

في هذه الصورة تتواجد قيم التردد الأساسي بين ١٠٠ و ١٥٠ هرتز، والصورة عبارة عن رسم طيفي لكلمة «مسافرون» بصوت ذكر المؤلف، ونلاحظ أن التردد الأساسي في المقطع الثاني أعلى منه في بقية المقاطع.



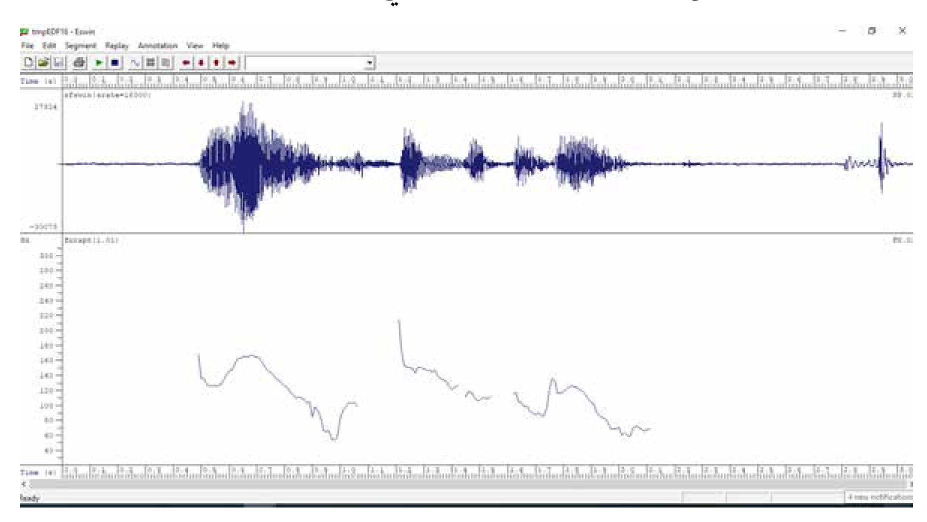
ويعرض الشكل الآتي منحنى التردد الأساسي لجملة استفهامية.



الشَّكل ١٨: منحنى التردد الأساسي لجملة استفهامية

يعرض (الشَّكل ١٨) مخططًا للتردد الأساسي لجملة استفهامية تمثل النغمة الصاعدة، حيث نلمح فيها أنها لم تحتوِ على ذلك الانخفاض الذي رأيناه في نهاية منحنى التردد الأساسي في الشكل السابق، والجملة (هل تصدق ذلك؟) بصوت المؤلف.

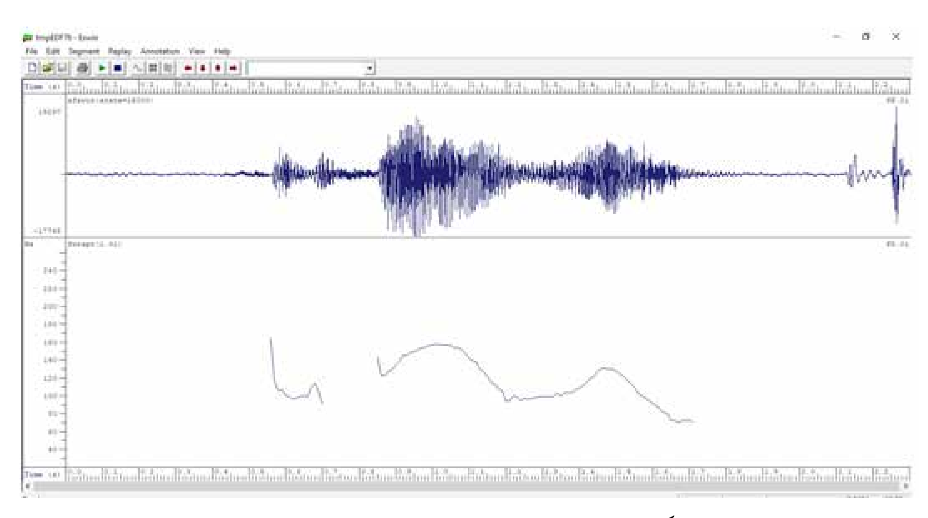
بينما يظهر (الشَّكل ١٩) منحنى التردد الأساسي لجملة خبرية.



الشَّكل ١٩: منحنى التردد الأساسي لجملة خبرية

ويعرض الشكل مخططًا للتردد الأساسي لجملة خبرية تمثل النغمة الهابطة؛ حيث يبدأ فيها التردد الأساسي منخفضًا ثم يزداد في منتصف الجملة ليصل إلى أقل انخفاض في نهاية الجملة، والجملة هي (نعم أصدق ذلك) وتمَّ تسجيلها بصوت المؤلف.

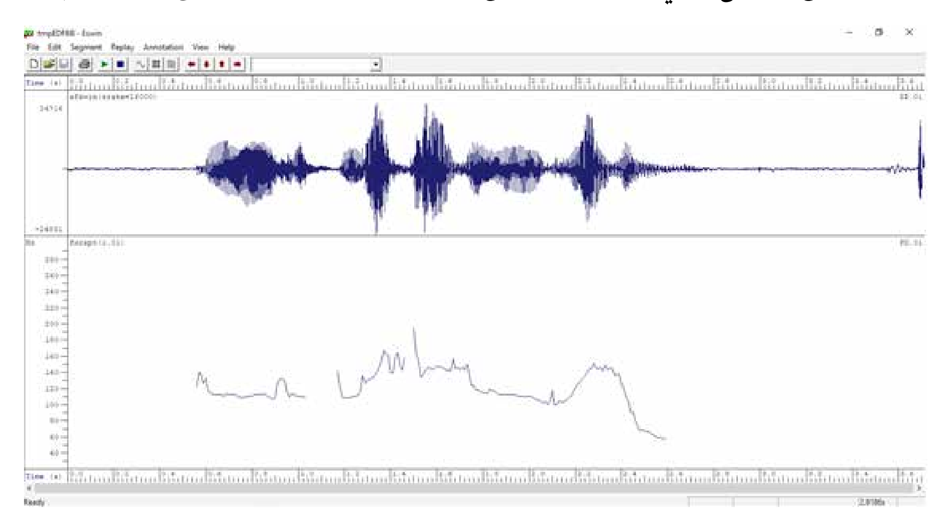
وقد دأب بعض اللغويين المعاصرين على الزعم بأنَّ هناك شكلًا ثالثًا لمنحنيي التنغيم السابقين - الاستفهامي والخبري- وهو المنحنى التنغيمي الخاص بالجملة التعجبية، إلا أنني قد قمت بجمع قاعدة بيانات صوتية لتسعين جملة تعجبية منطوقة من ثلاثة أشخاص، فلم أهتدِ إلى أي خلاف ظاهر بين طبيعة نطق هذه الجمل وبين الجمل الخبرية.



الشَّكل ٢٠: منحنى التردد الأساسي لجملة تعجبية

ويعرض (الشكل ٢٠) مخططًا لتردد الأساسي لجملة تعجبية، وقد ظهرت بنغمة هابطة تماثل النغمة الهابطة الملازمة للجمل الخبرية؛ حيث يبدأ فيها التردد الأساسي منخفضًا ثم يزداد في منتصف الجملة ليصل إلى أقل انخفاض في نهاية الجملة، والجملة هي (سبحان الله!).

ويعرض الشكل الآتي نموذجًا آخر من الجملة التعجبية المعتمدة على أداة تعجب.



الشَّكل ٢١: منحنى التردد الأساسي لجملة تعجبية معتمدة على أداة تعجب

يعرض هذا الشكل مخططًا للتردد الأساسي لجملة تعجبية معتمدة على أداة تعجب، ولم يظهر فيها أي خلاف عن الجمل الخبرية؛ حيث يبدأ فيها التردد الأساسي منخفضًا ثم يزداد في منتصف الجملة ليصل إلى أقل انخفاض في نهاية الجملة، والجملة هي (ما أصدقه من رجل!).

٤ , ٢ , ٣. الأداء الشخصي (Speaker Variability):

يتمثل الأداء الشخصي (Speaker Variability) في مجموعة من الفوارق الصوتية المرتبطة بأداء المتكلم، مثل: لجنس واللكنة والعمر وسرعة الأداء الكلامي من ناحية، وبوسيلة انتقال هذا الصوت سواء بطريقة مباشرة أو عبر الهواتف العادية أو المحمولة أو عبر الأثير الإذاعي من ناحية أخرى. إنَّ كل عنصر من هذه العناصر يضفي بظلاله على طبيعة هذا الصوت ويؤثر على عملية إدراكه وتمييزه. وتعد هذه النقطة من الأهمية بمكان في عمليات التعرف الآلي على الصوت المنطوق. والتي عادة ما تتطلب الاعتماد على عمليتين متوازيتين من التحليل الإحصائي للمتغيرات المتعددة للصوت، وهما تحليل المكون الرئيسي (PCA) وتحليل المكونات المستقلة (ICA).

وتمثل عملية التباين هذه إحدى الصعوبات الأساسية في تمييز الإشارات الصوتية، حيث إنها شديدة الارتباط ببعضها البعض، وهناك صعوبة أخرى في التعرف على الكلام، تتمثل في صعوبة التعرف على الإشارات الصوتية عند نقلها عبر الهواتف المحموعة أو عبر الأثير الإذاعي، حيث يتوقع حدوث التباس كبير بين بعض الأصوات المتشابهة مثل صوتي اليسن والشين. ولكي يتم تحليل هذه الإشارات الصوتية ينبغي الاعتماد على مجموعة من الأدوات الإحصائية المتاحة في هذا الصدد، مثل تحليل المكونات الرئيسية (principal component analysis) والتي تعرف اختصارًا بـ (PCA) وتحليل العناصر المستقلة (independent component analysis) والتي تعرف اختصارًا بـ (ICA) (H. Hotellings. 2005: 423) و (Hyvarinen, 2000: 418).

٤ . من التقنيات الصوتية الحاسوبية.

سبقت الإشارة إلى تعريف ابن جني للصوت اللغوي باعتباره «أصواتًا يعبر بها كل قوم عن أغراضهم». فالكلام هو الوسيلة الطبيعية للتواصل بين الجنس البشري. فلا يجد الناس أدنى مشقة تذكر في التواصل الصوتي حيث إنه لا يتطلب أي تدريب خاص. ومع انتشار الحاسوب واعتباره أحد متطلبات الحياة اليومية باتت الحاجة ملحة إلى التعامل معه كما يتعامل بعضنا مع البعض، وأصبح الناس يتطلعون إلى استخدام الكلام كوسيلة اتصال للتفاعل مع أجهزة الكمبيوتر، بدلًا من استخدام لوحات المفاتيح وأجهزة التأشير، وهو ما أدى إلى تسارع العمل في تقنيات إنتاج الكلام وتحليله والتعرف عليه (R. Arun, 2004: 14).

لقد أدى التطور التكنولوجي والحاسوبي إلى إحداث طفرة كبرى في تقنيات الصوت اللغوي، التي باتت تلعب دورًا أكبر في المجالات الأكاديمية والاقتصادرية والسياسية على حد سواء، ومن أبرز هذه التقنيات تقنية تحويل النص العربي المكتوب إلى صوت منطوق (Text To Speech) وتقنية التعرف الآلي على الصوت المنطوق (Speech Recognition) وتقنية البحث الصوتي (Audio Indexer) وتقنيات الترجمة الصوتية (Audio Translation) التي تعتمد على مزج تقنية تحويل النص المكتوب إلى صوت منطوق مع تقنية التعرف الآلي على الصوت المنطوق. وسوف نحاول الوقوف على تقنيتين من هذه التقنيات بمزيد من الإيضاح، هما: تقنية تحويل النص إلى صوت، وتقنية التعرف على الكلام المنطوق.

١ , ٤ . تقنية تحويل النص العربي المكتوب إلى صوت منطوق (TTS).

مع الطفرة التي أحدثتها التقنية في عالمنا المعاصر، وتزامنًا مع الثورة الإلكترونية التي صبغت كل مجالات الحياة أصبحت تقنيات معالجة الصوت البشري، وعلى رأسها تقنية تحويل النص العربي المكتوب إلى صوت منطوق text-to-speech مطلبًا مهما ليس فقط لتسهيل عمليات التواصل بين المتحدثين وإنما لأغراض اقتصادية وتجارية بل وتتعداه إلى جوانب سياسية وأيدولوجية، بالإضافة إلى الأبعاد الأكاديمية والاجتماعية التي تتمثل في مساعدة ذوي الاحتياجات الخاصة عن طريق تحويل الكتب والدوريات والجرائد والنصوص المكتوبة بشكل عام إلى منطوقة ليستفيد منها المكفوفون والأسوياء على حد سواء (الغامدي وآخرون، د.ت: ٧).

ويمكن القول: إنَّ تقنية تحويل النص العربي المكتوب إلى صوت منطوق عبارة عن تقنية حاسوبية تهدف إلى قراءة أي نص بصورة آلية، سواء تم تقديمه على شكل نص مكتوب بامتداد (text.*) أو (doc.*) أو (pdf.*) ... إلخ أو تمَّ تقديمه بعد مسحه ضوئيًّا بامتداد (JPEG.*) أو (BMP.*) أو (GIF.*) إلخ.. عن طريق الاستعانة بنظام التعرف الضوئي على الأحرف (OCR).

وبالتالي فهي تقنية مختلفة تمامًا عن أنظمة عرض الكلام المسجل عبر أجهزة الكاسيت أو الفيديو أو الكمبيوتر، حيث إنَّ الهدف هنا تخليق الكلام آليًّا وليس تخزين كلمات أو جمل محددة ثم عرضها وقت الحاجة.

وتجدر الإشارة هنا أن هذه التقنية تختلف كليًّا عن تطبيقات الاستجابة الصوتية التي يتمُّ استخدامها على شكل واسع في محطات القطار أو خدمة العملاء لدى شركات الاتصالات العالمية؛ حيث إنَّ ذلك النوع من آلات التحدث (المحددة سلفًا) التي يطلق عليها أنظمة الاستجابة الصوتية تنتج خطابًا اصطناعيًّا وليس تخليقيًا، وذلك عن طريق سلسة من الكلمات المعزولة (isolated words) أو أجزاء الجمل (parts of sentences). وغالبًا ما يتم الاعتماد على هذا النوع من الأنظمة عندما تكون المفردات المقصودة قليلة نسبية، وغالبًا لا تزيد عن ١٠٠ كلمة في سياق محدد سلفًا، أما في سياق تقنيات TTS فإنه من المستحيل تسجيل أو تخزين جميع كلمات اللغة، وبالتالي فمن الأنسب والأيسر تعريف «تقنية تحويل النص إلى كلام» على أنه تخليق أو إنتاج الكلام بواسطة الآلات، عن طريق تحويل هذا النص إلى صوت التلقائي (Thierry, 1997: 13).

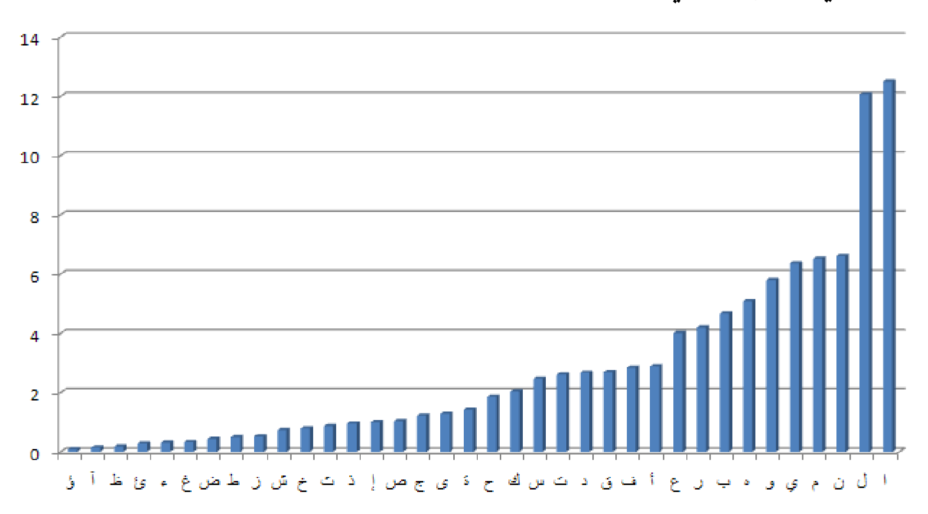
١ ,١ ,٤ . مراحل معالجة النص العربي وموائمته مع صوت منطوق.

أولًا: جمع المادة النصية.

يعد جمع المادة النصية المكتوبة أولى مراحل معالجة النص العربي في مشروع تحويل النص العربي المكتوب إلى صوت منطوق (TTS)، ولا بد أن تكون المادة النصية متوافقة مع الغرض من الناطق الآلي، بمعنى أنه إذا كان الهدف من إعداد الناطق الآلي قراءة الأخبار السياسية أو الرياضية أو الدينية فينبغي أن تكون كل قاعدة البيانات النصية مركزة على تلك الأخبار السياسية أو الرياضة أو الدينية، وإذا كان الغرض عامًّا بمعنى

أنه ناطق آلي لكل أنواع النصوص فينبغي أن تكون قاعدة البيانات النصية على نفس القدر من العمومية.

ولا يقتصر الأمر على نوع النصوص فقط، بل يجب مراعاة نِسَب توزيع الحروف والحركات العربية الواردة في النص؛ لتغطي كافة التراكيب الصوتية، بمعني أنه يجب أن تتواجد كل صوامت اللغة وحركاتها، وفي الوقت نفسه يجب مراعاة أن تكون نسب تردد هذه الصوامت والحركات موزعة توزيعًا عادلًا، وهذا التوزيع العادل لا يقتضي التساوي، فلا يُعْقل أن يتكرر صوت كثير التردد في النظام الصوتي العربي مثل صوت اللام أو الميم أو النون بنفس درجة تكرار صوت آخر قليل التردد في النظام الصوت العربي مثل صوت الضاد والغين والظاء. والشكل الآتي يوضح نسب توزيع الأصوات العربية في النص القرآني.



الشَّكل ٢٢: توزيع تردد إحصاءات الحروف العربية الواردة في النص القرآني، مرتبة على حسب تردد الصوت

يوضح الشكل السابق نسب ورود الأصوات العربية في القرآن الكريم؛ حيث ورد صوت اللام في القرآن ٦١٨٤٦٥ مرة بنسبة مئوية (١٢,٠٧ ٪)، وورد صوت النون ٣٣٨٦٤٦ مرة بنسبة مئوية (٦,٦١ ٪)، وجاء صوت الميم ٣٣٣٨٥٨ مرة بنسبة مئوية (٦,٥٢ ٪)، بينما لم يرد صوت الضاد إلا ٢٢٦٤٠ مرة بنسبة مئوية (٠,٤٤ ٪)، وصوت الغين ١٦٦٩٣ مرة بنسبة مئوية (٠,٣٣ ٪)، وصوت الظاء ٨٩٥٩ مرة فقط بنسبة مئوية (٠,١٨ ٪).

ثانيًا: المطابقة بين النص المكتوب والصوت المتوقع نطقه.

ويتم ذلك بتطبيق قاعدة كل ما يقرأ يكتب وما لا يقرأ لا يكتب، وذلك لسد الفجوة بين النظام الكتابي في اللغة العربية وبين النظام الصوتي، وقد أكَّد منصور الغامدي أن الباحثين العاملين في مجال النطق الآلي والتعرف الآلي على الكلام العربي يواجهون صعوبة إلى حد ما في التعامل مع النص العربي وذلك من حيث تحويل رموزه المكتوبة إلى رموز صوتية. ورغم أن العربية تكاد تكون أقدم لغة حُددت أصواتها ودون نظامها وقوانينها الفونولوجية قبل أكثر من ١٢ قرنًا إلا أنه لم توضع هذه الخصائص والقوانين الفونولوجية بشكل يمكن للعاملين في مجال الحوسبة الاستفادة منها بطريقة مباشرة (الغامدي وآخرون، د.ت: ٧)؛ حيث يُلزمنا نظام الكتابة العربي بإدراج حروف لا يتم نطقها وحذف حروف يجب نطقها، ومن أمثلة النوع الأول ما يلي:

- اللام الشمسية أينما وجدت، وهي لام التعريف التي تسبق الاسم النكرة فيصير معرفة، وتأتي مع ١٤ حرفًا، مجموعة في البيت التالي، وتمثل أوائل الكلمات:

  طب ثم صل رحما تفز زر ذا نعم --- دع سوء ظن زر شريفًا للكرم

  فأيما كلمة بدأت باللام الشمية نقوم بحذف هذه اللام من الكتابة الصوتية.

- الألف الفارقة، وهي الألف التي تلحق واو الجماعة في نحو: قاموا ، جلسوا ذهبوا، لأنَّ هذه الألفات لا تنطق، وإنما وضعت لعلل صرفية.
- الواو في كلمة عمرو لأنها لا تنطق وصلا أو وقفًا
- الواو الأولى في نحو: أولئك، وأولو، وأولي.

ومن أمثلة النوع الثاني ما يلي:

- إضافة واو ثانية في كلمة داوود. فتكتب: داوود.
- إضافة ألف في بعض أسماء الإشارة نحو: ذلك، وهذا، وهذان، وهذين، وهذه، وهؤلاء. فتكتب ذالك، وهاذا، وهاذان، وهاذين، وهاذه، وهاؤلاء.
- إضافة ألف في الأعلام مثل: الله، واللهم، وإله، والرحمن. فتكتب: اللاه، واللاهم، وإلاه، والرحمن.

- إضافة همزة متحركة مع ألفات الوصل عند البدء بها، لأن هذا البدء يترتب عليه نطق هذه الهمزة مثل همزة القطع، فمثلًا كلمة اسمع، تكتب هكذا: إِسْمَعْ.

ثالثًا: معالجة الاختصارات الكتابية.

المختصرات الكتابية في اللغة الإنجليزية عبارة عن «صورة مختصرة لكلمة word أو لاسم مركب compound noun أو لعبارة phrase ، تنشأ عن طريق ترك بعض حروف الكلمة أو استخدام أحرف معينة من كل كلمة، وعلى سبيل المثال فإنَّ "m" هي اختصار لكلمة "meter" في أحد التعيرات عن الطول، ويقابله في اللغة العربية الحرف ميم (م) الذي يدل على المعنى نفسه اختصارًا لكلمة (متر)، كأن يقال مثلا: طول الجدار ٣٠ م، كذلك فإنَّ المختصر (BBC) هي اختصار للعبارة British Broadcasting Corporation ، وقريب منها في العربية قولك (ش م م) اختصارًا لقولك شركة مساهمة مصرية، أو (س ح م) اختصارًا لعبارة سكك حديد مصر، وعادة ما يشيع في اللغات الحية استخدام بعض الاختصارات بصورة أكثر من الصورة الكاملة لهذا المختصر، ويتفاوت ذلك بين لغة وأخرى كثرة أو قلة (أحمد، ٢٠١٣: ٣٢٦).

ومع انتشار وسائل الإعلام المسموعة المرئية والمكتوبة، وتطور سبل الاتصال عبر الهواتف المحمولة والإنترنت أصبح استعمال المختصرات الكتابية ظاهرة تحتاج من القائمين على مشروعات تحويل النص العربي إلى صوت منطوق أن يولوا هذه الظاهرة الاهتمام اللائق بها؛ ومن ثم جمعها ثم إدراجها في أنظمتهم الناطقة.

ولا يخفى على كل متخصص أننا نستعمل في كتاباتنا بعض الرموز والمختصرات للدلالة على أمر ما يكوف معروفا عند من نخاطبهم، مثل: (ص.ب) التي نستعملها للدلالة على صندوق بريد، و (ه) التي نستخدمها للتعبير عن التاريخ الهجري، أو (م) التي نستخدمها للتعبير عن التاريخ الميلادي.

رابعًا: تشكيل النص آليًّا.

إنَّ وضع علامات التشكيل يؤدي إلى فك الالتباس الدلالي ووضوح المعنى، فعلى سبيل المثال إذا قمنا بتكوين كلمة ثلاثية من الجذر (ك ت ب)، فمن الممكن أن تكون كَتَبَ أو كُتِبَ أو كُتُبٌ، أو كَتْبٌ أو كَتَّبَ ... إلخ، وقد تقوم السليقة اللغوية بفك هذا

الالتباس تلقائيًّا عن طريق المعرفة اللغوية الثابتة في العقل الفردي، وإن كانت تلك المقدرة تختلف من قارئ إلى آخر، أما الحاسب الآلي فإنه لا بد من برمجته ليحمل المعرفة اللسانية التي عند المتحدث العربي ليصل إلى النتيجة نفسها، وهي عملية في غاية التعقيد. وقد بذلت مراكز علمية وبحثية عديدة جهدا مضنيا لتحقيق هذه الغاية إلا أن الهدف لم يتحقق بشكل مرض إلى الآن. ومن ثم كانت هناك محاولات لما يعرف بالتشكيل الجزئي diacritization partial؛ وهو تشكيل بعض الحروف وترك غيرها، كتشكيل الحروف المتصلة ببنية الكلمة وترك الحروف في نهايتها التي غالبا ما يكون لها علاقة بالنظام النحوي العربي أو تشكيل الكلمات الأكثر شيوعا وترك النادر منها (الغامدي وآخرون، د.ت: ٨).

وبدون تشكيل الحرف العربي لا يمكن لنظام نطق آلي معرفة التضعيف من عدمه والتنوين والصوائت التي تلي الصوامت، مما يجعل عملية النطق الآلي للنص الغير مشكل عملية مستحيلة (الغامدي وآخرون، د.ت: ٨).

خامسًا: قراءة النص بصوت واحد أو أكثر من القراء أو المذيعين الماهرين.

ونقصد بها مرحلة التسجيل وإعداد قاعدة البيانات الصوتية، وهي أهم مرحلة، حيث يرتبط مدى نجاح المشروع بجودة التسجيل وموافقته للمواصفات الفنية المعتمدة، ويجب أن نراعي النقاط الآتية في التسجيل:

- لا بدَّ أن تكون النصوص المختارة ذات تنوع فونيمي يتواءم مع نسب توزيع الأصوات في اللغة الفعلية.
- يتم الإحصاء على مستوى الفونيم الثلاثي وليس الثنائي أو المفرد.
- جودة تسجيل الصوت يجب ألا تقل عن ٣٢ك * ١٦ بت.
- التسجيل بصيغة استريو.
- تتزامن عملية تسجيل الصوت مع تسجيل الذبذبات الحنجرية (EGG).
- يكتفى بتسجيل قاعدة البيانات بصوت مذيع واحد.
- لا تقل المدة الصوتية الخالصة – بعد حذف الفترات الصامتة قبل قراءة كل جمل وبعدها – عن ثلاث ساعات.

- تتم عملية التسجيل داخل أستوديو محكم، وباستخدام جهاز لاقط صوت (مايك) على درجة عالية من الدقة، ومجهز بتقنية استبعاد الضوضاء المصاحبة للثوت، ويتم استخدام بطاقة صوتية معتمدة (Sound card).

سادسًا: تشكيل النص صوتيًّا، وفقًا للقراءة الفعلية.

في هذه المرحلة نقوم بإعادة النظر في تشكيل النص بعد قراءته من لدن المذيع أو القارئ الذي ستستخدم صوته في النظام، وهنا ينبغي عدم الاعتماد على القواعد النحوية المعيارية، بمعنى أنه إذا لحن المذيع أو القارئ الذي سنبني النظام على صوته في جملة مثل: شرح المدرسين الدرس، فإنَّ هذا الخطأ لن يترتب عليه أية مشكلة في إعداد قاعدة البيانات، لأن الاعتماد هنا على الوحدات الصوتية وليس على القواعد النحوية.

سابعًا: تخليق الكتابة الصوتية (Transcription).

ونقصد بها تحويل النص المكتوب باللغة العربية وفقًا لقواعد الأبجدية الصوتية الدولية (IPA) International Phonetic Alphabet أو وفقا لمنهج التقييم الفونيمي للأبجدية الصوتية (SAMPA)، وقد تم ابتكار النظام الأول (IPA) منذ أواخر القرن التاسع عشر، واستطاع الباحثون من خلاله ابتكار ألفبائية جديدة للتعبير عن أصوات اللغات المختلفة. وتحاول هذه الألفبائية الحفاظ على قرب العلاقة بين الرموز الكتابية والصّوت المنطوق. وكانت المراجعة الأخيرة لها في عام ١٩٨٩م. ويعمل نظام IPA على إيجاد رمز لكلّ فونيم في كلّ لغة بشريّة، ووضع علامات تشكيلية للتّغيّرات التي تعتري أية لغة من اللغات (MacMahon, ١٩٩٦: ٨٣٢).

وقد قسمت هذه الجمعية اللغوية الأصوات الإنسانية إلى ستة أقسام، هي:

- رموز لأصوات صامتة رئوية.
- رموز لأصوات صامتة غير رئوية.
- رموز لأصوات صائتة (الحركات).
- رموز لأصوات نغمية.
- رموز لعلامات تمييزية.
- رموز الفوقطعيات.

أما النظام الآخر وهو نظام (SAMPA) فيعد أكثر الأنظمة استخدامًا في برامج التحليل الصوتي الحاسوبي، وهو برنامج تم تطويره أواخر ثمانينيات القرن الماضي ليغطي ستة لغات أوروبية ضمن برنامج EEC ESPRIT للبحث والتطوير في مجال تكنولوجيا المعلومات. وفيه تمَّ الاعتماد على معظم الرموز الصوتية التي ابتكرتها IPA؛ ومن ثم إضافة علامات أخرى بديلة لبعض الأصوات لتكون أكثر ملائمة مع الحروف الحاسوبية، فمثلا تم استبدال الرمز [@] بالرمز [ə] الذي تم استخدامه من لدن IPA، ليعبر عن إحدى صوائت اللغة (L. DeMiller, 2000: 125).

وقد زاد من أهمية نظام (سامبا) ما تمتع به من التوافق وسهولة التعامل مع الحاسب الآلي؛ بحيث يجد كل رمز صوتي مكانًا له على لوحة المفاتيح بالجهاز، مما سهَّل استخدامه لدى الباحثين الحاوسوبيين. وقد تم إدراج الرموز الصوتية العربية في هذا النظام على النحو المبين في الجدول الآتي:

المجموعة الأولى: الحركات:

الرمز	المثال	الكتابة الصوتية
I	ظل	D`il
A	حل	X\al
U	عمر	umr`?
i:	عيد	i:d`?
a:	مال	ma:l
u:	فول	fu:l

المجموعة الثانية: أنصاف الحركات:

الرمز	المثال	الكتابة الصوتية
W	واحد	wa:hid
J	يوم	Jawm

المجموعة الثالثة: الصوامت:

وتتفرع عنها صوامت انفجارية، وصوامت احتكاكية، وصوامت أنفية، وصامت تكراري، وصامتان جانبيان، على النحو المبين أدناه:

الصوامت الانفجارية:

الرمز	المثال	الكتابة الصوتية
B	باب	ba:b
T	تسع	`?tis
D	دار	da:r
`t	طابع	`?t`a:bi
`d	ضرب	d`arab
K	كبير	kabi:r
G	جميل	gami:l

الصوامت الاحتكاكية:

الرمز	الكلمة المثال	الكتابة الصوتية
F	فيل	fi:l
V	نفين	nivi:n
T	ثلاث	Tala:T
D	ذكر	Dakar
`D	ظلام	D`ala:m
S	سعيد	sa?`i:d
Z	زميل	zami:l
`s	صغير	s`aGi:r
S	شمس	Sams
Z	جميل	Zami:l
X	خطاب	xit`a:b
G	غرب	Garb

الرمز	الكلمة المثال	الكتابة الصوتية
\X	حلم	X\ilm
(\?) `?	علم	alam`?
H	هواء	?:hawa

الصوامت الأنفية:

الرمز	المثال	الكتابة الصوتية
M	مال	ma:l
N	نور	nu:r

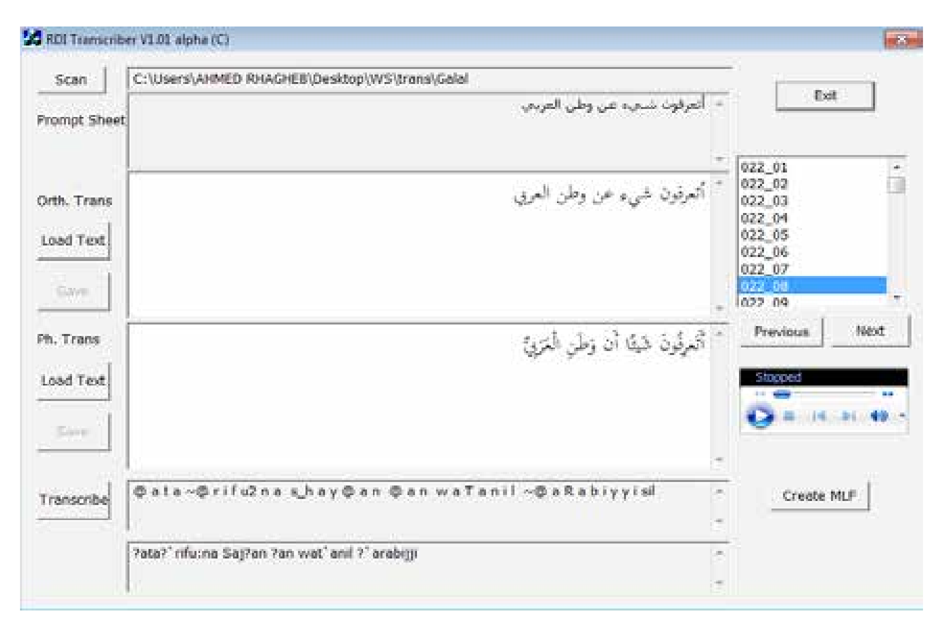
الصوت التكراري:

الرمز	الكلمة المثال	الكتابة الصوتية
R	رمال	rima:l

الصوتان الجانبيان:

الرمز	الكلمة المثال	الكتابة الصوتية
L	لا	:la
`l	الله	al`l`ah?

ولا يعني ذلك بالضرورة أنه يجب على الباحث في المجال الصوتي الحاسوبي أن يقوم بتحويل كل حرف أو حركة إلى رموز الكتابة الصوتية السابقة بطريقة يدوية، بل هناك الكثير من الأدوات الحاسوبية التي تحوِّل النص العربي الـمُشَكَّل إلى أحد أنظمة الكتابة الصوتية بطريقة آلية، على النحو الموضح في الشكل الآتي:



الشَّكل ٢٣: برنامج محول الكتابة الصوتية (RDI Transcribe V1.01 Alpha) الذي ابتكرته الشركة الهندسية لتطوير النظم الرقمية (RDI)

ثامنًا: موائمة الجمل النصية مع التدفقات الصوتية.

حيث نقوم في هذه المرحلة بإعداد مجموعة من الملفات لكل جملة منطوقة، وهذه الملفات هي:

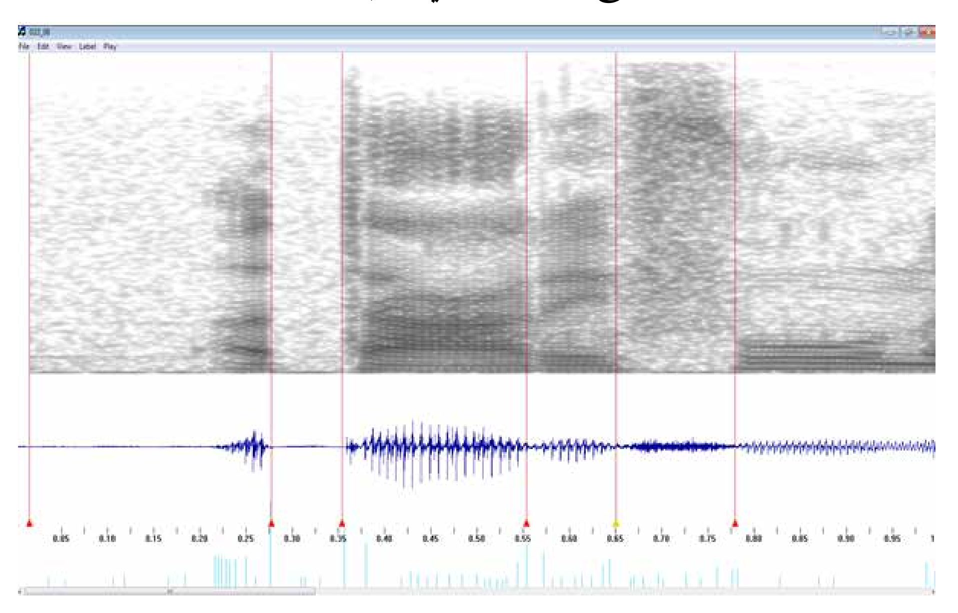
- ملفات الصوت: بامتداد (Wave)، ويمثل المادة الأساسية للمشروع.
- ملفات النصوص: بامتداد (Text) ، ونكتب فيها النص الموافقة للصوت المنطوق.
- ملفات الكتابة الصوتية: (Transcription)، متوافقا مع إحدى الأبجديات الدولية.
- ملفات الأنو: (Ano) لتحليل خصائص كل فونيم (الخصائص الكيفية).
- ملفات اللاب:(Lab)، لتحديد زمن كل فونيم (الخصائص الكمية).

تاسعًا: استنباط بداية ونهايات الفونيمات آليًّا.

يتم تحديد بدايات ونهايات كل فونيم في كل كلمة من كلمات قاعدة البيانات الصوتية عن طريق تمريرها على تقنية (HMMs) عن طريق تحديد الخواص الأكوستية لكل فونيم، والتي تتمثل في: الجهر أو الهمس، الشدة أو التوسط أو الرخاوة، الاستعلاء أو الاستفال، الإذلاق أو الإصمات ...إلخ.

عاشرًا: تحديد بدايات ونهايات الفونيمات.

ويتم هذا التحديد أولًا بشكل آلي عبر (HMMs) ثم يتم ضبط بدايات ونهايات كل فونيم من لدن خبراء صوتيين، وهناك الكثير من الأدوات والبرامج التي تسهم في إنجاز هذه المهمة، ومنها برنامج التحديد الزمني الآتي.



الشَّكل ٢٤: برنامج التحديد الزمني (Segment Lab) الذي أنتجته الشركة الهندسية لتطوير النظم الرقمية (RDI)

وبذلك تكون قاعدة البيانات صالحة لدمجها في نظام تحويل النص العربي المكتوب إلى صوت منطوق.

٢ , ٤ . تقنية التعرف على الكلام المنطوق (ASR).

تعد تقنية التعرف على الكلام المنطوق واحدة من أهم فروع تخصصات اللسانيات الحاسوبية التي تهدف إلى تطويع المناهج والتقنيات لتكون أكثر قدرة على التعرف على اللغة المنطوقة وترجمتها أو تحويلها إلى نصوص إلكترونية (Lee, Chin-Hui, 199: 232).

كما يعرف هذا التخصص أيضًا باسم التعرّف التلقائي على الكلام (automatic speech recognition) ويشار إليه اختصاراً ب (ASR) أو تقنية تحويل الصوت المنطوق إلى نص مكتوب (speech to text) والتي يشار إليها اختصارًا بـ (STT). إنه نوع من العلوم يمكنه دمج المعرفة والبحوث اللغوية مع علوم الحاسب داخل إطار مجال الهندسة الكهربائية.

ومما يميز تلك التقنية اعتمادها بشكل أساسي على «تدريب النظام» (training) أو (enrollment) عن طريق قاعدة بيانات صوتية وعنونتها بناءً على معطيات التحليل الصوتي الحاسوبي، وغالبًا ما يتم الاعتماد على عدد كبير من المتحدثين للغة أو للهجة الواحدة، شريطة أن تختلف فئاتهم العمرية وخواصهم الصوتية، فيقرأ كل متحدث منهم بشكل فردي نصًا لغويًا، وربما يتم الاكتفاء بقراءة مجموعة من المفردات المعزولة السياق، وذلك وفقًا للغرض من إنشاء النظام سواء أكان الغرض طرح النظام لشريحة كبيرة من المتحدثين أو قصره على فرد واحد حيث يكتفي النظام بتحليل صوت هذا الشخص ويستخدمه من أجل التعرف على الصوت نفسه لاحقًا، ويتسم هذا النظام بالدقة الشديدة، وتسمى هذه الطريقة التابعة للمتحدث "speaker dependent" في حين تسمى الطريقة غير المعتمدة على المتحدث(speaker independent Lee, Chin-Hui, 1991: 232) و (Martin, E. A. 1987: 778).

وهناك الكثير من التطبيقات المتاحة التي تقدم خدماتها اعتمادًا على تقنية التعرف الآلي على الصوت المنطوق، ومنها على سبيل المثال التطبيقات المثبتة في الهواتف النقالة أو المحمولات الإلكترونية بشكيل عام، حيث تتيح هذه التطبيقات إعطاء مجموعة من الأوامر الصوتية كفتح برنامج ما على الجهاز أو الاتصال بشخص من قائمة المراسلات أو إرسال بطاقة معايدة أو تهنئة أو ما إلى ذلك.

ومنها أيضًا ما يتم تركيبه في أجهزة التحكم عن بعد، وعن طريقها يمكنك مثلًا فتح جهاز التكييف وضبط درجة حرارته عن طريق الأوامر الصوتية، ومن ذلك أيضًا التحكم في توجيه الكراسي المتحركة التي يستخدمهـــا ذوو الاحتياجـــات الخاصة (P. Nguyen 2010: 127).

كما يتم استخدامها بشكل أوسع في الكتب الإلكترونية التي يمكنك من خلالها إعطاء أوامر محددة للتطبيق فيستجيب لها كتقليب صفحات الكتاب أو قراءة فصل من فصوله، أو التحكم في حجم الخط ونوعه ولونه ....إلخ.

والأهم من ذلك كله استخدام تلك التقنية وإدارجها في تقنية مشابهة لها، وهي البحث الصوتي الآلي (Speech Audio Indexer) التي تتيح للمتصفح العثور على مواضع معينة من ملفات صوتية مطولة (Lori Lamel, 2008: 8).

ببليوجرافيا مرجعيَّة.

١. ابن جني، أبو الفتح عثمان. (١٩١٣)، الخصائص، تحقيق محمد على النجار، دار الكتب المصرية، الطبعة الرابعة، الجزء الأول، الطبعة الأولى.

٢. ابن جني، أبو الفتح عثمان. (١٩٥٤) سر صناعة الإعراب، تحقيق مصطفى السقا وآخرون، دار الكتب العلمية، بيروت، لبنان، الجزء الأول، الطبعة الأولى.

٣. أحمد، أحمد راغب. (٢٠٠٢) دور المؤثرات السياقية في تحديد المدى الزمني للفونيم، مجلة الدراسات اللغوية والأدبية، السنة الثالثة، العدد الأول.

٤. أحمد، أحمد راغب. (٢٠٠٤)، فونولوجيا القرآن: دراسة لأحكام التجويد في ضوء علم الأصوات الحديث، رسالة ماجستير غير منشورة، كلية الآداب جامعة عين شمس.

٥. أحمد، أحمد راغب. (٢٠١٣) المختصرات الكتابية في اللغة العربية - دراسة استقرائية تحليلية. مجلة الدراسات العربية والإسلامية، دار العلوم جامعة المنيا، العدد ٢٧.

٦. أحمد، راغب أحمد. (٢٠٠٩) نظام التعرف الآلي على الصوت القرآني، حفص- دراسة توثيقية اختبارية، المؤتمر الدولي الأول للتعلم والتعليم عن بعد، الرياض.

٧. أنيس، إبراهيم. (١٩٦١). الأصوات اللغوية. القاهرة: دار النهضة العربية.

٨. أيوب، عبد الرحمن. (١٩٨٤): الكلام إنتاجه وتحليله، مطبوعات جامعة الكويت.

٩. البهنساوي، حسام. (٢٠٠٤)، علم الأصوات، مكتبة الثقافة الدينية، القاهرة، مصر، ط١.

١٠. حسان، تمام. (١٩٩٠)، مناهج البحث في اللغة العربية ، مكتبة الأنجلو المصرية.

١١. الحمد، غانم قدوري. (٢٠٠٧)، الدراسات الصوتية عند علماء التجويد، دار عمار للنشر والتوزيع، عمَّان، الطبعة الثانية.

١٢. حميداني، عيسى واضح. (٢٠١٠). الصوت اللغوي دراسة وظيفية تشريحية، دار غيداء للنشر والتوزيع، الطبعة الأولى.

١٣. العطية، خليل إبراهيم، (١٩٨٣)، في البحث الصوتي عند العرب، منشورات دار الجاحظ للنشر، بغداد، الطبعة الأولى.

١٤. عطية، سليمان. (٢٠١٥)، الفونيمات فوق التركيبية في القرآن الكريم (المقطع -النبر -التنغيم) الأكاديمية الحديثة للكتاب الجامعي.

١٥. عمر، أحمد مختار. (١٩٩٧)، دراسة الصوت اللغوي، القاهرة، عالم الكتب.

١٦. الغامدي، منصور بن محمد. (٢٠٠٠)، الصوتيات العربية، مكتبة التوبة، الطبعة الأولى.

١٧. الغامدي، منصور بن محمد. (د. ت)، حسني المحتسب، مصطفى الشافعي. قوانين الفونولوجيا العربية، منشورات مدينة الملك عبد العزيز للعلوم والتقنية.

١٨. محمد، سهام. (٢٠٠٩) الفونيمات غير التركيبية (الثانوية): التنغيم أنموذجاً، بحث غير منشور، الجامعة الأردنية، كلية الآداب.

١٩. الموسوي، مناف. (١٩٩٨)، علم الأصوات اللغوية، عالم الكتب، بيروت، الطبعة الأولى.

٢٠. هلال، عبد الغفار حامد. (١٩٨٨)، أصوات اللغة العربية، مكتبة الأنجلو المصرية، ط٢.

٢١. الوعر، مازن. (١٩٨٩)، دراسات لسانية تطبيقية، دار طلاس، دمشق، الطبعة الأولى.

1. Bannon, Mike; Kaputa, Frank. "*The Newton–Laplace Equation and Speed of Sound*". Thermal Jackets. Retrieved 30 May 2019.
2. Lee, Chin-Hui, Rabiner, R. Lawrence and Pieraccini, Roberto (1991) '*Speaker Independent Continuous Speech Recognition Using Continuous Density Hidden Markov Models*'. NATO ASI Series, Vol. F75.

3. Lori Lamel, Jean-Luc Gauvain, *Speech Processing for Audio Indexing, International Conference on Natural Language Processing*, GoTAL 2008, Advances in Natural Language Processing pp 4-15
4. Martin, E. A. (1987), 'A *Two-stage Isolated-word Recognition System using Discriminant Analysis*', MIT Lincoln Laboratory, Technical Report 773
5. P. Nguyen (2010). *Automatic classification of speaker characteristics, International Conference on Communications and Electronics* 2010, INSPEC Accession Number: 11698809
6. H. Hotellings, "*Analysis of a complex of statistical variables into principle components*", J. Educ. Psychol., 24, pp.417-441, 498-520, 1933.
7. Huang, C., Chen, T., Li, S.Z., Chang, E., & Zhou, J. (2001). *Analysis of speaker variability. INTERSPEECH*.
8. Hyvarinen and E. Oja,(2000) "*Independent component analysis: algorithms and application*" Neural Networks 13. pp.411~430.
9. JL Flanagan, *Speech Analysis, Synthesis and Perception, Springer-Verlag*, New York, 1972
10. Jump up to:a b Blust, Robert. (2013). *The Austronesian Languages* (Rev. ed.). Australian National University.
11. Khattab, Ghada & Al-Tamimi, Jalal. (2014). *Geminate timing in Lebanese Arabic: The relationship between phonetic timing and phonological structure*. Laboratory Phonology, 5 (2), 231-269.
12. L. DeMiller, Anna & Rettig, James (2000). *Linguistics: A Guide to the Reference Literature* (2nd ed.). Libraries Unlimited. ISBN 1-56308-619-0.
13. L. E. Baum, A *statistical estimation procedure for probabilistic functions of Markov processes*, IDA-CRD Working Paper No. 13

14. MacMahon, Michael K. C. (1996). "*Phonetic Notation*". In P. T. Daniels and W. Bright (eds.). The World's Writing Systems. New York: Oxford University Press. pp. 821–846. ISBN 0-19-507993-0.
15. Mr. R. Arun Thilak & Mrs. R. Madharaci (2004), "*Speech Recognizer for Tamil Language*", Tamil Internet, Singapore.
16. Stevens, K.N.(2000), *Acoustic Phonetics*, MIT Press, ISBN 0-262-69250-3, 978-0-262-69250-2
17. Thierry Dutoit, *An Introduction to Text-to-Speech Synthesis, Springer*; 1997 edition (April 30, 1997), P 13
18. Uszkoreit, Hans. "*DFKI-LT - What is Language Technology*". Retrieved 16 November 2018.
19. Wall, Joan (1989). *International Phonetic Alphabet for Singers: A Manual for English and Foreign Language Diction. Pst*. ISBN 1-877761-50-8.
20. http://hyperphysics.phy-astr.gsu.edu/hbase/Sound/intens.html , Retrieved 30 May 2019.
21. http://www.intellaren.com/articles/ar/a-study-of-arabic-letter-frequency-analysis, Retrieved 30 May 2019.
22. http://www.wataonline.net/site/modules/newbb/viewtopic.php?topic_id=1777, Retrieved 30 May 2019.
23. https://www.cooleditpro.com. , Retrieved 30 May 2019.
24. https://www.goldwave.com/goldwave.php., Retrieved 30 May 2019.
25. https://www.grc.nasa.gov/WWW/BGH/sound.html, Retrieved 30 May 2019.
26. https://www.phon.ucl.ac.uk/resource/sfs/ , Retrieved 30 May 2019.

الفصل الرَّابع

الذَّكاء الاصطناعيّ وتعليم اللُّغة العربيَّة

نحو مِنَصَّةٍ تعليميَّةٍ مُتَكَامِلَة

د. نعيم محمد عبد الغني

ملخص

هذا الفصلُ ليسَ تأصيلا وتأريخًا لعلاقةِ الذكاءِ الاصناعيّ باللغة العربية، وإنما هو مشروعٌ طموحٌ لمنصةٍ علميَّةٍ تتكاملُ فيها التقنياتُ التي تعالجُ اللغاتِ الطبيعيةَ في منصَّةٍ واحدةٍ تتاحُ للجميعِ على اختلافِ مستوياتِهم وأعمارِهم. وقد سلَّطْتُ في هذه الدراسَةِ الضوءَ على أهمّ التَّقْنِياتِ ذات التَّطبيقات الـمُتعدِّدة، مع التَّركيز على البُعد التَّعليميّ.

وقد أبنتُ عن كيفيَّة توظيف هذه التَّطبيقات في تعليم العربيَّة، كما أبنتُ عن جوانبها الإيجابيَّة والسَّلبيَّة، وكيف يمكنُ تطويرُها؛ لتحققَ النتيجةَ المرجوةَ من توظيفِها في بيئةِ التَّعليم.

وانطلقَ الفصلُ بتسلسل منطقي يُعَرِّفُ اللغةَ ويرصدُ حاجةَ الإنسانِ لها، ووسائل تعليمِها التقليدية وغير التقليدية، ثم يرصدُ تعليمَ اللغة في مستوياتِها: الصوتية والصرفية والتركيبية والدلالية، ثم عرضتُ نموذجا مُصَوَّرًا لمشروعِ مِنَصَّةٍ تعليميَّةٍ مُتكَامِلَة.

الكلمات المفتاحية:

اللغة العربية، الحاسوب، التعرف على الصوت، التعرف على الصورة، تكنولوجيا التعليم.

١ . اللُّغة والتَّعلُّم.

اللغةُ أكثرُ ما يميزُّ الإنسانَ عن باقي المخلوقاتِ، فالوظيفةُ الأساسية المميِّزة لعقل الإنسان هي مقدرتُه على إنتاج أنظمةٍ رَمْزِيَّةٍ والتمكنُ من استعمالِها، وعلى رأس هذه الأنظمة الرمزيَّةِ النظامُ اللغويُّ المستعملِ في التواصلِ وتمثيلِ المعلومةِ وتخزينِ المعرفةِ ونقلها.

إن اللغةَ وِعاءُ حضارةِ الإنسان وفكرُه، فمن خلالِها يستطيعُ التعبيرَ عن احتياجاته، ومن ثم فهو بشكل تلقائيٍّ يسعى لاكتسابِ لغتِه من محيطِه الذي يعيشُ فيه بداية من أمِّه وأسرتِه وانتهاءً بمجتمعِه الذي يتفاوتُ ضِيقا وسِعة حسْبَ سعي الإنسانِ واحتكاكِه بالثقافات الأخرى، ذلك الاحتكاكُ الذي قد يدفعه لتعلُّمِ لُغَةٍ أخرى غيرِ لغتِه الأم.

إذًا فالإنسانُ عنده دافعٌ لِتَعلُّمِ اللغة؛ كي يستطيعَ أن يتواصلَ مع محيطِه الذي يعيشُ فيه، غير أن هذا التعلمَ في حدِّ ذاتِه عمليةٌ مركبةٌ وصعبة، تستدعي من المتعلمِ توظيفَ معارفَ ذهنيةٍ وآلياتٍ نفسيةٍ واجتماعيةٍ. وهذا العبءُ لا يحس به إلا المعلم النظامي، أما من يُعلِّمُ اللغة بشكل فطري كالأم التي تعلمُ ابنها عن طريق الاحتكاكِ المباشر بها فإنها لا تضعُ خطة منظمة، بل يكون التعليم بشكل فطري سلسل ويكونُ محدودا بطبيعة تفاعل الأم والأسرة مع هذا الطفل، أما المعلمُ النظامي للغة فإن عليه عبئًا في تعليم اللغة؛ لأنَّه مطالَب بتوظيفِ معارِفه ومهارَاتِه ليُعَلِّمَ الطلابَ، بدايةً من التخطيطِ للمنهجِ، ثم التدريس والتواصل بين طلابه وانتهاءً بالاختبارات والتقييم.

بهذا المدخل يتبين لنا أن الإنسان يمر بمرحلتين في تعلم اللغة، الأولى: هي المرحلة العشوائية الطبيعيةُ التي يكتسبُ فيها الإنسان لغتَه من محيطِه، والمرحلةُ الثانية تتمثل في التعليم النظامي للغة.

واللغةُ كأداةِ اتصالٍ تتكونُ من مستوياتٍ عدَّة: تبدأ بالصوْتِ ثمَّ الصرفِ ثم التركيبِ وأخيرا الدلالة، ومتعلّمُ اللغةِ ينبغي أن يتعلمَ هذه المستوياتِ بالتوازي، فيتعلمُ أصواتَ الحروفِ ويُفَرِّقُ بينَ الحركاتِ والصوامتِ ثمَّ يُكَوِّنُ ثَرْوَةً لُغويَّةً من المفرداتِ يستطيعُ أن يُكَوِّنَ منها جُملا صحيحةً تركيبيًّا ولها مَعْنَى دِلالِيّ يَفْهَمُه المُخَاطَب. والمحصلة النهائية أن يكون مُتَعَلّم اللغةِ قَادرًا على استخدامِها تحدُّثًا وكتابةً وقِراءةً وسَماعًا، وَهَذِه العَمليةُ هي

الغايةُ التي يَسْعَى إليها مُتَعَلّم أيةِ لغةٍ، ومنها اللغةُ العربيةُ، بِغَضِّ النَّظَرِ عَمّا تمتازُ به العربيَّة من خصوصياتٍ تُمَيِّزُها عن باقي اللغات، لكنَّ هذه العمليةَ معقدةٌ؛ حيثُ تتطلبُ بحثًا فيسيولوجيًّا وفيزيائيًّا لتوضيحِ طريقةِ إنتاجِ الكلام، ثُمَّ بَحْثًا لُغُويًّا في دراسةِ خصائصِ هذا الكلام، فنحن اللغويين - في واقع الأمر - لا ندرسُ طريقةَ إنتاج الكلام، بلْ نَدْرُسُ الكلامَ بَعْدَ إنتاجِه، فنحللُ خصائصَه الصوتيةَ والصرفيةَ والنحويّةَ والدلاليّة. والذكاءُ الاصطناعيُّ أمامَه إذًا تحدٍ كبير، الأول: في كيفيةِ إنتاجِ الكلامِ ومحاكاةِ العقلِ البشريِّ في ذلك، والثاني في كيفيَّةِ الـمُحافَظَةِ على خصائصِ الكلامِ وقواعدِه في كلِّ لغة.

٢ . طبيعة إنتاج اللغة.

تتخذ التعريفاتُ المختلفة للغة عِدَّةَ مُنْطَلَقاتٍ تَصِلُ غالبُها إلى نتيجةٍ واحدةٍ، هي أن اللغةَ أداةٌ للتواصلِ الإنسانيِّ، فَهِي أصواتٌ يُعَبِّرُ بها كُلُّ قَوْمٍ عن أغراضِهم (ابن جني،١/ ٣٣) أو هي نظامٌ من الرموزِ الاعتباطيَّةِ يتمُّ بواسطتِها التعارفُ بينَ أفرادِ المجتمع، وتخضعُ هذهِ الأصواتُ للوصفِ من حيث المخارجُ أو الحركاتُ التي يقومُ بها جهازُ النُّطق ومن حيث الصِّفاتُ والظواهرُ الصوتيَّةُ الـمُصاحبةُ لهذا النطق (عمايرة، ١٩٨٧: ١٥).

وإذا كانت اللغةُ أصواتًا أو إشاراتٍ متفارقةً - على حد تعبير دي سوسير- فهذا يعني أنها رموزٌ ووحداتٌ لغويةٌ تعملُ ضِمْنَ نظامٍ يقومُ بوظيفةٍ اتصاليةٍ بين أفرادِ المجتمع. وَهذِه الرمُوز يخزنها العقلُ البشري؛ ليُنتجَ منها أصواتا وكلامًا له معنى ودلالة وفق آليَّةٍ مُعَقَّدَة، هذه الآليةُ تَنَاوَلَهَا الدَّارِسُونَ على مُستويِّن، المستوى الأولُ هو إنتاجُ الكلامِ في المخ، والمستوى الثاني فيسيولوجيةُ النُّطْقِ. فأما إنتاجُ الكلامِ في الدِّماغِ فيكونُ بتفعيلِ تَرْمِيزَاتٍ لغويّةٍ للتحدثِ والفهمِ والقراءةِ والكتابةِ، فالمتحدث يبدأ في التفكير في معنى ما يريد أن يتحدث به ثُمَّ يَخْتَارُ له المفرداتِ المناسبةَ، وهذه المفرداتُ لها أصواتٌ وشكلٌ صرفيٌّ، ثُمَّ بَعْدَها يُرَكِّبُ الجُمْلَةَ في رأسِه لينْطِقَها في نهايةِ المطافِ، هذه العمليةُ تترجمُ إلى حركةِ الفَمِ والفكين واللسانِ والحنكِ والحنجرة، وتتمُّ في الإنسانِ الطبيعيّ بشكل سريع.

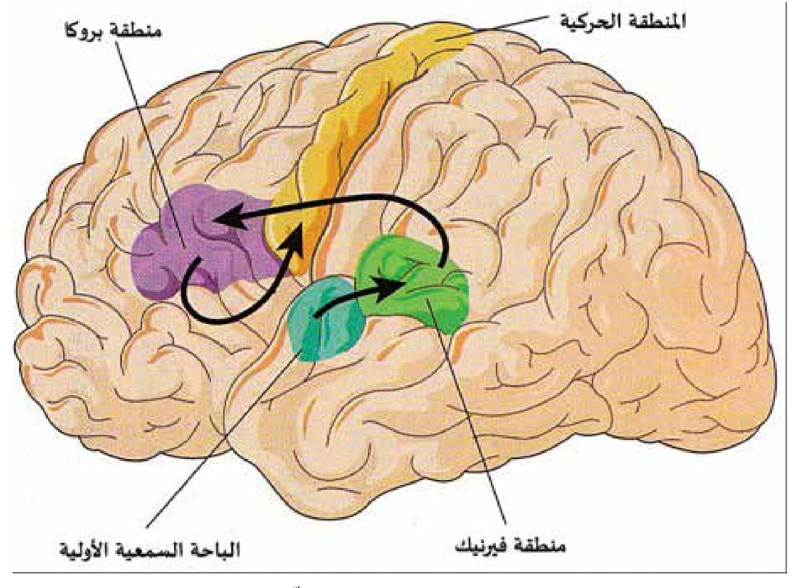
إنَّ آليةَ إنتاجِ الكلامِ واستقبالِه هيَ حالةٌ من حالاتِ الإعجازِ في خلقِ البشر، حالةٌ وقفَ علماءُ اللغةِ والطبِّ أمامَها كثيرًا، فعَالِمُ النَّفسِ فيجُوتِسْكِي ذَكر أنَّ اللغة تتم من

خلالِ الاتصالِ الخارجِي مع البشرِ، وترتيب الأفكار في العقلِ، فهناك نظامان يعملان: نظامٌ داخليٌّ، وآخرُ خارجيٌّ، وكلا النّظامين يستخدمانِ شفرة لغويّة واحدة (جوديث جرين، ١٩٩٢: ١٣ - ١٥).

أما تشومسكي فإِنَّهُ يرى أن إنتاجَ اللغة يَمُرُّ من خلالِ الكفاءةِ والأداءِ، فكُلُّ إنسانٍ يحتوي دماغُه على معرفةٍ لُغويَّةٍ يتم تغذيتُها من محيطِ مجتمَعِه، وَبَعْدَ إضافَةِ الـمُدْخَلات اللغويةِ يُصْبِحُ الإنسانُ قادرًا على التكلم (Chomsky, ٢٠٠٠: ٧٧-٧٩).

إن هذه المسألةَ درَسَها باستفاضةٍ علمُ اللغة العصبيّ (Neuro linguistique)، ولهُ أفرُعٌ متعددةٌ يهُمنا منها الآن ما يتصل بطريقةِ إنتاجِ الكلام، ففي تشريح الدماغ - كما في (الشكل ١) هناك مناطق مسؤولة عن اللغة هي: منطقة بُروكا واللُّحَاء السمعي واللُّحاء الحركيّ الرئيسيّ ومنطقة فيرينك، وعند نُطْقِ أَيَّةِ كلمة مقْروءة يجبُ أن تصلَ المعلوماتُ أولا إلى القشرة البصريةِ التي ترسلُ بدورِها المعلوماتِ إلى مِنْطَقَةِ الكلامِ الخلفيةِ بما فيها منطقةُ فيرينكِ، ثم تنتقل المعلوماتُ إلى مِنْطَقَةِ بروكا، ثُمَّ إلى اللُّحَاءِ الرئيسيِّ.

إنَّ هذِه المناطقَ مسؤولةٌ عن السمعِ والنُّطقِ، فهي أدوات الاتصال المباشرة، وأيّ خلل فيها يُحْدِثُ صممًا أو فقْدًا للنطقِ أو على الأقل اضطرابٌ فيهما (شهداء، ٢٠١٥، ٨٣-٨٩).



الشكل ١: مناطق إنتاج الكلام في الدِّماغ البشري

٣. الذَّكاء الاصطناعيّ وإنتاج اللغة.

والسؤالُ الآن: هل يستطيعُ الذَّكَاءُ الاصْطِنَاعيُّ مُحَاكاةَ هذا النظام المُعَقَّدِ في تركيبِه وطريقَةِ أدائِه؟ وما الذي تَوَصَّلَ إليه حتى الآن؟ وهَلْ يُمكنُ للآلَةِ أنْ تَحُلَّ محلَّ الـمُعلِّم في تعليمِ اللُّغة؟

والجوابُ نعم؛ لأن بين اللغة والذكاءِ الاصطناعيّ عاملا مشتركا يتمثل في الترميز، فإذا كان العقل البشري يحتاج إلى أن نُعرِّفَه على شكل الحروف التي هي رموز في الأساس، فبالإمكان أن ندرب الحاسوب على قدر كبير من هذه الرموز، فالعلاقة بين اللغة والحاسوب تأخذ طابعا تبادليًّا، والالتقاء بين اللغةِ والحاسوبِ أمرٌ حتميٌّ لأن الإنسانَ محورُ النشاطِ اللُّغويّ، والإنسانُ هو الذي منحه الله القدرة على تصميم الحاسوبِ وتشغيله، وهذا يمكّنُه من صناعة لغات البرمجة التي تخدّم لغته التي يتكلم بها (علي، ١٩٨٨: ١١٣، ١١٥).

إنَّ صناعةَ تكنولوجيا المعلوماتِ تقوم على الانطلاق من اكتشافِ حاجات البشرية؛ لِتَدْرُسَ إمكانية تحقيقِها في صورةِ تطبيقاتٍ سهلةِ التناول، ومن حاجاتِ البشريةِ الملحةِ الآن ضرورةُ التواصلِ اللُّغويِّ، سواء أكان هذا التواصلُ بين أبناءِ اللغةِ الواحدةِ، أم عن طريقِ الترجمةِ بينَ لُغاتٍ متعددةٍ، وفي الأمرين يتدخلُ الذَّكَاءُ الاصطناعيُّ الذي ما زالَ يُوَاجِهُ تحدياتٍ كبيرةً حتى الآن، ويبلغُ هذا التحدي ذروتَهُ في التَرْجَمَةِ الآليةِ التي مَا زَالتْ حُلْمًا يُراودُ علماءَ التكنولوجيا الذين عَكَفُوا الآن على دِرَاسَةِ إمْكَانِيَّةِ تَحْقِيقِه؛ فبدأوا بدراسةِ خَصَائِصِ اللغاتِ في أصواتِها ومفرداتِها وتراكيبِها لمعرفة المعاني التي ستكونُ مُتَرْجَمَةً بِدِقَّة.

وتتم هذه الدراسات عَبْرَ خُبَرَاء في هندسة التكنولوجيا وخُبَرَاء في اللغات؛ حيثُ شُغِلُوا الآن في حَصْرِ الموارِدِ المُتَاحَةِ في محاولةٍ لحلِّ بعضِ المشكلاتِ التي سُتُواجِهُهُم في التَّرْجَمَةِ الآلية؛ رغبةً في تواصلِ البشرِ في كافة أنحاءِ المعمورةِ بشكلٍ كاملٍ مقروءٍ ومسموعٍ، فالكلُّ يَرْجُو أَنْ يَطَّلعَ بِسْهُولَةٍ على الكُتُبِ والصُّحُفِ المُدَوَّنَةِ بغيرِ لُغَتِه التي يتكلمُ بها، ويريدُ أن يفهمَ محادثةً فوريةً بجهازٍ بسيطٍ عِنْدَمَا يتكلمُ معَ أحدٍ من غيرِ أبناءِ لغتِه، كَمَا يَوَدُّ أنْ يُوَصِّلَ مَا عِنْدَهُ للآخرين الذينَ يتكلمونَ بلسانٍ غير لسانِه، والكلُّ

يتساءلُ: هلْ سيأتي اليوم الذي أتكلم فيه بالهاتف بلغتي ليُتَرْجَم كلامي عَبْرَ لغة أخرى فيفهمُها الأجنبي، ويأتي ردُّه الذي هو بلغةٍ أخرى مُتَرْجَمًا للغتي في سرعةٍ ودقة؟!

وكل هذه الأماني يَنظرُ إليها الباحثون باهتمام فيتطلَّعون لحلِّها، فيدفعهم أملهم للعمل، فتكونُ البدايةُ في تحليلِ اللغات ومعرفةِ مشكلاتِها والقدرِ المشتركِ بين اللغاتِ الذي من الممكن أن يكونَ قاعدةً لهذه الترجمةِ؛ ليوفرَ الجُهْدَ والوقتَ. وحتى لا تكونَ البدايةُ من فراغ، فقد بدأَ الغربُ في البحثِ عن خصائصِ اللغاتِ ووضعوا مناهجَ مختلفةً للبحث كان لها دورُها في صناعةِ تكنولوجيا عربية؛ فعلى سبيل المثالِ عندما بَدَأَ الإنجليزُ في محاولةِ صُنع تطبيقاتٍ تكنولوجيَّةٍ على اللغةِ الإنجليزيةِ انطلقوا من الصوتِ لِصُنْعِ برنامجٍ لتعليمِ اللغةِ الإنجليزيةِ للناطقين بغيرها، ويعتمدُ البرنامجُ على تقنية التعرّفِ على الصَّوْت، فقاموا بصنع برنامج يُسَمَّى (Tell me more) الذي يُعلِّم كيفية نطقِ اللغة الإنجليزيةِ بصورة صحيحة، وإذا استقام النطقُ للمتعلمِ استطاع بسهولةٍ أن يكتسبَ اللغةُ بمفردَاتِها وتركيبِها.

أما الترجمةُ الآليةُ فنجحتْ في جرِّ القائمين في تكنولوجيا المعلومات إلى آفاقٍ بحثيةٍ أخرى لا بُدَّ من ارتيادِها لخدمةِ البشريَّة؛ مما خلقَ رؤيةً واضحةً إلى حدٍّ كبيرٍ للخارطَةِ البحثيةِ في عالمِ التكنولوجيا، عن طريق تحديدِ الأهدافِ، ثُمَّ تعيينِ المتطلباتِ الضروريةِ؛ لتحقيق هذه الأهداف التي ستُسْتَخْدَمُ في تطبيقات مفيدة للبشريَّة، ومن يُطالع مثلا موقعَ منظمةِ (ELRA) - وهي منظمةٌ تهتم بدراسةِ اللغاتِ الأوربية - يرى أنَّهم وضعوا خُطَّةً بدأت من ٢٠٠٣ وحتى ٢٠١١، وحدَّدوا هدفهم النهائي في ٢٠١١ الماضي والمتمثلُ في الترجمة الآلية للغة الهولندية، وما زالت خُطَّتُهم تسيرُ بنجاحٍ حتى الآن، بعد أن حقَّقُوا جُزْءًا كَبيرًا من هدفهم بالنسبة للغةِ الهولندية.

٤ . الذكاء الاصطناعيّ واللغة العربيَّة.

إذًا بعد أن تعرفنا على الخارطة البحثية للذكاء الاصطناعيِّ وتعاملِه مع اللغاتِ بشكل عام، فمن الممكن استثمارُ هذا النجاحِ في اللغاتِ الأخرى للتعاملِ مع اللغةِ العربيةِ، وخاصةً في مجالِ التعليم، لكنَّ الأمرَ لَيسَ سهْلًا؛ فهناكَ مُعْضِلاتٌ كُبْرَى خَاصةٌ باللغةِ العربيةِ نفسِها، وبطريقةِ التفكيرِ في إنتاج الكلام واستقباله، فهل يمكنُ جعلِ الحاسوبِ مُحاكيًا للعقلِ البشريّ في تعليم اللغة العربيَّة بتعقيداتِه التي أشرنَا إليها؟

إن اللغةَ العربية تمتاز بخصائص فريدة تساعدُها على برمجتِها آليًّا، وبشكلٍ يندرُ وجوده في لغاتٍ أخرى، فالانتظامُ الصوتيُّ في اللغةِ العربيةِ والعلاقة الوثيقة بين طريقة كتابتِها ونطقِها يدلُّ على قابلية اللغة العربية للمعالجة الآلية بشكل عام، وتوليد الكلام وتمييزه آليا بصورة خاصة (علي، ١٩٨٨: ١٧٤). واللغةُ العربيةُ وُصِفَت بأنَّها لغةَ جبرية؛ لشدةِ انتظام كثير من خصائصِها الصوتية والصرفية والإعرابية، وهذا يؤكد قابلية العربية لأن تختزل في البرمجة الآلية من خلال معادلات رياضية (علي، ١٩٨٧: ٧٥٣، ٧٥٤).

وفي العقودِ الثلاثة الماضيةِ كانت المحاولاتُ جادةً في اللغةِ العربيةِ؛ لجعلِ الحاسوبِ مُحاكيًّا للعقلِ البشريِّ، من خلالِ العملِ في مستوياتِ اللغةِ العربيَّةِ كإطارٍ منهجيٍّ، حيثُ كانت هناك جهودٌ صوتيةٌ وصرفيةٌ ونحويةٌ ومُعْجَميةٌ ودلاليةٌ، ونِسْبَةُ النجاحِ تفاوتتْ حَسْبَ طبيعةِ كُلِّ مستوى، فالنجاحُ على مستوى الصوت والصرف كان مُرْضِيًّا، بينما على مستوى التركيبِ والدلالةِ كانت النتائجُ غَير مُرضية، وإن أمكنَ الاستفادة منها؛ وهذا ما سيظهرُ في الصفحاتِ التاليةِ.

٥. الذكاء الاصطناعيّ وتعلُّم النطق.

تعلُّمُ النُّطقِ من خلال الذكاءِ الاصطناعيِّ يكونُ عبر تقنية «التعرف على الصوت»، وهي تقنيةٌ لها تطبيقاتٌ كثيرةٌ، ونلْحَظُها في الهواتفِ المحمولةِ؛ حيثُ القراءةُ الآليةُ للأسماءِ والرسائلِ وغيرِها، وهذه التقنيةُ يمكنُ استثمارُها في تعليمِ نُطْقِ أصواتِ الُّلغةِ بشكلٍ عامٍ واللغةِ العربيةِ بشكلٍ خاص.

من الأمثلة على ذلك تقنية «حفص» لتعليمِ أحكام التجويد. وتتوافقُ فكرةُ هذه التقنية مع بحث للدكتور صلاح حامد؛ حيث استطاعَ بطريقةٍ رياضيةٍ توظيفَ التقنياتِ لمعرفةِ أخطاءِ التلاوة (حامد، ٢٠٠٤)، وسأركزُ على دراسةِ حامد في هذا المقام؛ لأنِّي كُنتُ أحدَ الذين عملوا مع الباحث في هذه التقنيةِ التي تُعلّمُ نطقَ أصواتِ العربيةِ، من خلال مرحلتين:

المرحلة الأولى: بناء نموذجٍ معياريٍّ لأصواتِ اللغةِ العربيةِ، ويشملُ هذا النموذجُ الفونيمات الأساسيَّةَ من الصوامتِ والحركاتِ والفونيماتِ فوق التركيبيةِ من النبرِ والتنغيم.

أما المرحلة الثانية فإنها تشملُ جمع احتمالات الخطإ عند المتعلمين في الشريحة المستهدفة، فإذا كنَّا نستهدفُ الأطفالَ فإننا نحاولُ جمعَ احتمالاتِ الخطإ عندهم وتحليلَ أسبابِه وإعطاءَ النموذجِ الأمثل للنطق الصحيح.

وما يحدث بعد هاتين المرحلتينِ هو أننا ندَّربُ الحاسوبَ على معيارية النطقِ الصحيح للفونيم فنعطيه مثلا أن الباءَ تتكونُ من فونيمين مرققين (Ba) فمن أتى بخلاف ذلك النطق يوجهه الحاسوب إلى وجود خطإ، ويمَكِّنُه من التَّكْرَار بشكلٍ سليمٍ حتى يخرجَ لنا هذا النطق الصحيح، وإذا قلنا إن الطاءَ حرفٌ مفخَمٌ فإننا نبني نموذجا معياريًّا صحيحًا للطاءِ الـمُفَخَّمَة؛ بحيث لا يمكنُ قبولُ أي صوتٍ يُخالِفُ هذا النموذجَ المعياريَ، فهذه مرحلةُ بناءِ النموذج، لكنَّ هذه المرحلةَ بها عدةُ مُشْكِلات لم تتمكن التكنولوجيا حتى الآن أن تَحُلَّها، من ذلك مثلًا:

أولا: الفونيمات الأساسية، يمكن ضبطها في الغالب - خاصة في مجال التفخيم والترقيق- لكن الصوائت أو الحركات تختلف أطوالها حَسْبَ أداء المتكلم، وهذا يتضح جليًّا في قراءةِ القرآن الكريم، فمن الممكن أن تقرأَ (بسم الله الرحمن الرحيم) في أزمنة مختلفة تتفاوتُ في أطوالِها حَسْبَ أداءِ بطيءٍ وسريع وكُلُّها صحيحةٌ، والحاسوبُ يَضْبِطُ أطوالَ المدودِ بعمليةٍ رياضيَّةٍ بَحْتَة، فإذا افترضنا أن طولَ الألفِ في قراءةِ «الحَدْرِ» مثلا يساوي (X) فإن طولَه أيضا في قراءة «التحقيق» يُساوي خمسةَ أضعاف، وهذا أيضا يختلفُ في الكلام العاديِّ عند اختلافِ اللهجاتِ وتنوعها.

ثانيا: أما الفونيمات التركيبية، وهي النبر والتنغيم فإنَها أيضا تمثل مشكلة؛ لأنها لا تعتمد على الأداء الصوتي فقط، وإنما تعتمد على تلوينِ الصَّوتِ الذي يختلفُ بشكلٍ كبيرٍ من شخص لآخر، كما تعتمدُ على لغة الجسد وما يَعْتَرِيه من انفعالاتٍ أثناءَ الكلام، وهذا ما يُحاول الذكاء الاصطناعيُّ حلَّه حتى الآن.

أما المرحلة الثانية فإنها تمثلُ جمع الأخطاءِ الـمُحْتَمَلَة من الشريحة المستهدفة، وهذا الجمعُ يتم من خلالِ توزيعِ جُملٍ ونصوصٍ تتم قراءتُها من الشريحةِ التي تُصَمَّمُ لها التَّقنية، ثم بعد ذلك يتم تحليلُها، فنذكر أن احتمالاتِ الخطإ في الطاءِ أن تُنطق تاءً مثلا، ويكثر هذا عند الإناث، أو أن تنطقَ مُشَرَّبةً بضاد لتكون طاء شديدة، كما في بعض لهجات

الصعيد بجنوب مصر. وهكذا نتتبع احتمالاتِ الخطإ وأسبابَه في العينة التي نُدرِّبُ الحاسوب عليها، وكلما كانت العينة كبيرةً صارت الدِّقَّةُ عاليةً في اكتشاف الأخطاء.

بهذا المفهوم أنتجت شركة (RDI) تطبيقات لتعليم أصوات العربية، ومنها برنامج (حفص) المشار إليه، والذي يمكن اختصار الكلام حوله على النحو التالي:

- يستخدم هذا البرنامج -كما أسلفنا - تقنية (التعرف على الصوت)، وبدأت بتعليم أحكام التجويد من خلال جمل مختارة من القرآن الكريم، تتمثل في كلمة وكلمتين حتى خمس كلمات أو ست كلمات.
- يبدأُ النموذج المعياريُّ في نطق الكلمة حسب القواعد المعمول بها.
- يستمعُ المتعلم جيدا، ويحاولُ ترديد ما سمع فإن استطاع تقليد ما سمع بشكل صحيح؛ لتكون فونيمات المتدرب مطابقةً لفونيمات النموذج، حينها تخرجُ له رِسالةٌ تفيده بأنَّه ُعلى صواب، وإنْ أخطأَ تُحَدِّدُ له بشكلٍ بصريٍّ مكانَ الخطإ، ثم يستمعُ إلى رسالةٍ تُفيدُ أنَّه أبدلَ الثاء سِينا مثلا أو جعلَ المدَّ سبعَ حركات والصوابُ أن يكونَ سِتًا وهكذا..

إن الأمثلةَ التي اعتمد عليها برنامجُ «حفص» مُخْتَارةٌ بعناية شديدة، وهي تطبيقات على دروس التجويد بتصنيفاتها التقليديةِ في الكتب، ومصحوبةٌ بشرحٍ للمصطلحات والقواعدِ الصوتية لنطق القرآن الكريم، حَسْبَ مستوى المتعلم. وهذه التقنيةُ كي تصل إلى مستوى يناسب متدربين من كافة المستويات احتاجت لتحليلِ الأخطاء عند شريحة كبيرة من المجتمع العربي على تنوع لَهَجَاتِه واختلافِ ثقافاته ومستوياته الاجتماعية وأعمار أفراده، فكان تَجْرِبَةً رائدةً بدأت بواكيرُها في مطلع القرن الحادي والعشرين، ولا يزالُ التطويرُ فيها مستمرا.

لقد وفر البرنامج بيئة تعليمية تحاكي المعلم التقليدي، لكن بطبيعة الحال لا غنى عن المعلم التقليدي، وتبقى التقنيةُ وسيلةً مساعدةً تُحَقِّقُ نتائجَ جيدةً، وبتجربَة حفص على فصلين دراسيين من الطلاب أُجْرِيَ لهما اختبار تحديد مستوى، ثم كان لفصل منهما تدريس حكم التفخيمِ والترقيق بالطريقة التقليدية، والفصل الثاني تَعَلَّم الحكمَ من خلال برنامج حفص، ثم أُجري للفصلين اختبار، وكانت النتيجة أن أداءَ المتعلمين عن

طريقِ التكنولوجيا أكثرَ نجاعة، وربما يرجعُ ذلك لأن التكنولوجيا توفرُ وسائلَ بصرية وسمعية للمتعلم بخلاف الدارس بشكلٍ تقليديٍّ؛ لأنه يعتمدُ على مهاراتِ الأستاذ وما يبتكرُه من وسائلِ تعليم - غالبا - ما تكون تقليدية.

ويمكن استثمارُ هذه التقنية وتوسيعُها لتعليم أصوات العربيةِ بشكل عام، وخاصة في سن الطفولةِ، فالتعليم الصحيحُ للغة يبدأ من الأصواتِ التي تساعدُ على الكتابة بشكلٍ صحيح، كما يمكن تطبيقُ هذه التقنيةُ في تعليم اللغةِ العربيةِ للناطقين بغيرها، وتعليم ذوي الاحتياجات الخاصةِ من الذين يعانون مشكلاتٍ في النطقِ، فالوسائلُ الإثرائيةُ الـمُصاحبةُ من الأنشطةِ والأسئلةِ والمسابقاتِ وغيرِها من مُحَفِّزَاتِ التعليمِ تكونُ قيمةً مُضَافَةً لهذه التقنية الأساسية في تعليمِ النطقِ الصحيح.

إن الموجود من برامجِ تعليمِ أصواتِ اللغةِ الآن لا تعدو أن تكونَ حُروفًا يتعرف المتعلمُ أشكالها منفردةً ومتصلةً في كلماتٍ يُكَرِّرُها الـمُتَعَلّم، لكنهُ لا يتلقى تغذيةً راجعةً لطريقةِ النطق الصحيح، ومن ثَمَّ فتقنيةُ «التعرف على الصوت» تُكْمِلُ هذا النقصَ وتجعلُ الآلةَ أكثرَ تَفَاعُلًا مع المتعلم، لتنشأَ بذلك بيئةٌ أقربُ إلى الطبيعية.

إن الذكاء الاصطناعيَّ كأداة مُسَاعِدَة في تعليمِ أصواتِ اللغة العربيةِ أثبت نجاعة ونجاحًا، فهناك دراسةٌ أُجْرِيتْ بِجَامعة اليرموكِ على طلاب الصف الثاني الإعدادي، وعددهم أربعون طالبا وطالبة، قُسِّموا إلى مجموعتين متكافئتين، مجموعةٍ تجريبيةٍ وأخرى ضابطةٍ، واستمرت الدراسةُ ثمانيةَ أسابيع بواقعِ حصَّةٍ كل أسبوع، وتعلَّمَ الطلابُ في هذه الفترة سورتي الصافات وص، ثم أُجْرِيَ للمجموعتين اختبارٌ يقيس مستوى الأداء في التلاوة، فتفوقت المجموعة التجريبية التي استخدمت التكنولوجيا السمعية والبصرية على المجموعة الضابطة التي درست بشكل تقليدي (عبد الله، ١٩٩٠: ٦٢، ٦٣).

وهذه النتيجة أشرنا إليها في التجربة التي حاول فيها الدكتور صلاح حامد أن يتعرفَ فيها على أداءِ تقنية حفص في تعليم أحكام التجويد التي تُعَدُّ في صلب الدرس الصوتي للغة العربية، والتَجْرِبَتَانِ متباعدتان في الزمان والمكان والأهداف، لكنهما تثبتانِ شيئا واحدًا هو جدوى استخدام التكنولوجيا في تعليم النطق الصحيح.

٦ . برامج إثراء الثروة اللفظية.

في العموم تكون المعاجم اللغوية هي المصدر الذي يَحوي الألفاظ والشواهد النثرية والشعرية، وهذه المعاجم يُلحظ فيها أنها تعتمد على بعضِها بشكل كبير، وبها ألفاظ مهجورة غير مُسْتَعْمَلةٍ، أو بها ألفاظٌ تَصْعب على الدارسين. ولا يُمكن أن نطلبَ من المتعلم أن يحفظَ مُعْجَمًا كي يتمكن من اللغة، بل إننا نكسِبُه ثروةً لُغويةٍ من خلال تطبيقٍ عملي لممارسةِ اللغة تَحدُّثًا وكِتَابةً وقراءةً، وفي هذه الحالة يُكَوِّنُ المتعلمُ مُعْجَمًا خاصًا به يكتسبُ فيه المفرداتِ بطريقةٍ تراكُميَّة، وهذه هي الطريقةُ التقليديةُ في تنمية الثروةِ الُّلغوية، ويفيدنا الذكاء الاصطناعيُّ بتحديد أكثرِ المفرداتِ والأساليبِ اللغوية انتشارًا، ومن خلال هذا الحصرِ يُمْكِنُنا بناءُ منهجٍ تعليميٍّ تتفاوتُ مستوياتُه حَسْبَ الـمُتَلَقِّي؛ ومن ثَمَّ فإن الذكاء الاصناعي بإمكانه أَن يبنيَ مُعْجَمًا حديثًا لِتَتَبُّعِ الألفاظ والأساليب الأكثر استعمالا في اللغة العربية، وكيفية استخدام هذه الألفاظِ والأساليبِ في سياقاتِها الـمُختلفة، فالحاسوب يُحَدِّدُ جِذْرَ الكلمة آليا ووسائلَ اشتقاقِها والمواطنَ الـمُخْتَلفة التي تُسْتَخْدَمُ فيها الكلمة.

إن هذا المنهج له جذور ترجع إلى العالم الأمريكي إدوارد ثورنديك (١٨٧٤ - ١٩٤٩) الذي سعى إلى إيجاد وسيلة لتعلم اللغة الإنجليزية فألف كتاب «قائمة المفردات للمعلمين" Teacher's Wordbook ونُشر عام ١٩٢١م؛ وفيه استخلص أكثر الكلمات استخداما وشيوعا بناء على مدونة نصية مؤلفة من واحد وأربعين كتابا أبرزها: الإنجيل والأساطير الأمريكية والصحف والكتب المدرسية. (السعيد، ٢٠١٦).

وبعد هذه المحاولة الجادة كانت هناك محاولات مهمة مثل مشروع معجم كوبيلد الإنجليزي التعليمي (Cobuild English Dictionary)، وهو معجم يعتمد في بنيته الأساسية على مدونة لغوية إلكترونية، وفي هذا المعجم حصرٌ للمفردات والتراكيب الشائعة واستعمالاتها اللغوية، واعتمد المعجم على مدونة من عشرين مليون كلمة جمعت من مصادر مختلفة، وخرج المعجم بعد هذا الجهد الكبير إلى النور عام ١٩٨٧م، ثم أعيدت طباعته في عام ٢٠٠٠م بعد زيادة حجم المدونة إلى مئتي مليون كلمة، (السعيد، ٢٠٠٧: ٦٩ وما بعدها).

وهناك دراسة فتحي يونس عام ١٩٧٤ عن الكلمات الشائعة عند تلاميذ الصفوف الأولى من المرحلة الابتدائية (يونس، ١٩٧٤)، والمفردات الشائعة في اللغة العربية لداوود عبده عام ١٩٧٩م (عبده، ١٩٧٩). ومن الدراسات أيضا التي اعتمدت رصد المفردات الشائعة دراسة بعنوان «تصميم وتطوير مدوَّنة لغويَّة للعربيَّة المعاصرة a Corpus of contemporary Arabic» للباحثة لطيفة السليطي وهي دراسة ماجستير بجامعة ليدز البريطانية أنجزت عام ٢٠٠٤ (Al-Suliti, 2004).

وسأتوقف قليلا عند دراسة السليطي التي اعتمدت فيها على مدونة من ٤١٥ مقالة تقارب ٩٠٠ ألف كلمة، وشملت مجالات السياسة والاقتصاد والرياضة والإعلام، وبعد فهرستها وتحليلها استطاعت استخلاص أكثر ألف كلمة مكررة في اللغة العربية، وكان لفظ الجلالة الأكثر ترددا في هذه المدونة، ثم قامت بإحصاء أكثر ٥٠٠ متلازمة في اللغة العربية ونعني بالمتلازم أن تستدعي الكلمةُ كلمةً أخرى بشكل عفوي، مثل مكة المكرمة، المدينة المنورة، وقد أطلق عليها الدكتور تمام حسان مصطلح التوارد وأطلق عليها غيره التصاحبات اللفظية (عبد الغني، ٢٠٠٥: ٣٥).

وبناء على هذا المفهوم استخرجت لطيفة من مدونتها ٥٠ متلازمة، مثل: القرآن الكريم، العالم الإسلامي، الصحة العالمية... إلخ، وصنفت الباحثة متلازماتها إلى متلازمات ثنائية الكلمة أو ثلاثية أو أكثر من ذلك، فمثلا من المتلازمات الثلاثية ثاني أكسيد الكربون، ومن المتلازمات الرباعية: حقوق الإنسان في الإسلام.

ولم تكتف السليطي بهذا بل استخرجت من مدونتها إحصاء بأكثر مئة مسكوك أو تعبير اصطلاحي؛ ويقصد به التعبير اللغوي الذي يضم أكثر من من وحدة لغوية تستعمل باطراد في اللغة، وهذا المسكوك تكون له دلالة ثابتة تختلف عن الدلالة المعجمية للمفردة (عمر، ١٩٨٥: ٢٢، ٢٣) مثل: (العلاقات العامة، على قلب رجل واحد، بغض النظر عن، على سبيل المثال... إلخ)، ومثل هذه المسكوكات تثري معرفة الطالب وتجعله قادرا على التحدث بطلاقة وتجعل لديه رشاقة في الإنشاء وإنتاج النص الكتابي.

كما أحصت السليطي نسبَ استخدام الروابط والأدوات في المدونة، فكانت حروف الجر أعلاها ثم بعدها أسماء الإشارة، ثم الأسماء الموصولة.

إن مثل هذا العمل يمكن استثماره في إعداد منهج لإثراء الثروة اللفظية لكل مراحل التعليم، سواء لتعليم العربية لأهلها أو للناطقين بغيرها؛ لأن الاعتماد على المستعمل من الكلام يشعر الطالب بأهمية ما يدرسه، والذكاء الاصطناعي يمكن أن يساعدنا في بناء مدونة نستخرج منها المشهور من الكلام سواء في المفردات والجمل والتراكيب، ويساعدنا على تنقية المناهج من الحشو الذي لا يستفيد منه الطالب، وفي هذا المقام هناك عدة دراسات قام بها الدكتور المعتز بالله السعيد تمثل بشكل عملي كيفية توظيف المدونات اللغوية في تعليم العربية وتطوير مقرراتها التعليمية. ومن ذلك (مدونة اللغة العربية لأغراض تعليمية) التي تناولَ فيها ثلاثة مستويات مؤثرة في المهارات اللُّغوية الأربعة (الاستماع والتحدُّث والقراءة والكتابة). وتمثلت المستويات الثلاثة في: (الحروف، والكلمات، والتراكيب)، وخلص الباحث إلى أن الأكثر شيوعا يكون الخطأ فيه قليلا فمثلا تلاميذ المرحلة الابتدائية لا يجدون مشكلة في (ا، ل، ي، و، م) لأنها مألوفة لديهم في كلمات كثيرة تأتي أمامهم، بينما يخطئون في كتابة: (ؤ، ئ، ظ)، لقلة ورودها أمام هؤلاء الطلاب (السعيد، ٢٠١٦: ٥٧-٩١).

إن توظيف مثل هذه المدونات في منصة علمية سيعطي لنا نتائج إيجابية في تعليم اللغة العربية للجميع. وإذا اخترنا من المشهور مسكوكات ومتلازمات فإن هذا يمكن استثماره في تعليم اللغة العربية للناطقين بغيرها، فالمسكوكات أو التعابير الاصطلاحية التي تختلف في دلالاتها عن الموجود في المعجم يمكن أن تجعل المتعلم يقارن بين مسكوكات لغته ومسكوكات العربية، فتنشأ عملية احتكاك أو تفاعل إيجابي، وقد تكون هذه بداية لمعجم مسكوكات لغوي يشمل مسكوكات اللغة العربية وما يقابلها من مسكوكات في اللغات الأخرى.

إن بناء المدونات وصناعة المعاجم بهذه الصورة تحتاج إلى جهد مؤسسي، ولا يمكن للأفراد أن ينهضوا بذلك. لقد حاولت شركة صخر إنجاز معجم حديث للغة العربية احتوى على ٣٦ ألف مدخل بثمان وأربعين ألف كلمة مع عشرة آلاف تركيب مصاحب، وذلك مع سماتها النحوية والدلالية بمرادفاتها وأضدادها ومشتقاتها وسياقاتها المختلفة وترجماتها للإنجليزية وإمكانية نطقها بالعربية.

لقد عملت شركة صخر على بناء ذخيرة لغوية هائلة اعتمدت فيها على مدونة لكتابات معاصرة تبدأ من خمسينات القرن الماضي بالإضافة إلى مئة وعشرين صحيفة عربية وضمت المدونة أيضا البرامج الوثائقية المرئية التي أتيح محتواها الكتابي على الإنترنت، ولم تغفل المدونة مصادر اللغة الأساسية من القرآن والحديث والشعر. وبعد هذا الجمع تأتي مرحلة المعالجة التي تتطلب وجود محلل صرفي يعقبه مشكل آلي ليُبنى عليه بعد ذلك الناطق الآلي ثم المترجم الآلي، وهذه خطوات تسلم كل واحدة إلى الأخرى فلا يمكن البدء بالترجمة الآلية مثلا لأنها مبنية على ما سبقها من منجزات.

لكن هذه التجربة لم تكتمل حتى الآن، فأنماط نصوص اللغة العربية كثيرة جدا، وتحتاج إلى فريق كبير من الباحثين الذين يعملون على تطويرها وتوسيع المدونة ووضع قوانين للعلاقات بين المفردات، وهذا لا يمكن أن تنهض به مؤسسة خاصة، بل يحتاج لمؤسسة عامة تدعم رسميا وشعبيا وتكون لها صفة الاستمرارية والاستدامة.

إن هذه المدونة لا بد أن تعرض على المجتمع ويبدي رأيه فيها وتكون له صلاحية في إبداء الملاحظات التي يمكن الاستفادة منها في تحديث المدونة وتصحيح المعجم، وهناك تطبيق لفتَ انتباهي في هذا المقام، وهو تطبيق الرديف، الذي يهتم بالبحث عن معنى الكلمة ومفردها وجمعها وما يذكر لها من شواهد شعرية إن وجدت، ويتيح للمستخدم أن يقترح كلمات تضاف إلى قاعدة بياناته، لكن هذا التطبيق بسيط ولا يعتمد على مدونة نصية، بل اعتماده على ما ذُكِر في المعاجم، ولذلك فهو وإن سهل البحث عن بعض الكلمات فإنه يظل محدودا. فعدد مترادفاته تزيد قليلا عن خمسة آلاف مترادف، وبه ما يقرب من تسعة آلاف ضد، وجمع شواهد شعرية للكلمات التي يبحث عنها بالمترادف أو الضد ما يقرب من ثلاثة وثلاثين ألف كلمة، ورغم هذه المحدودية أجريت عليه أكثر من مليون عملية بحثية، وهذا يؤكد لنا أن أية مبادرة جادة في هذا المجال ستجد لها قراءً ودارسين.



الشكل ٢: تطبيق «الرديف» لتنمية الثروة اللفظية

ومن التطبيقات المفيدة في مجال إثراء الثروة اللفظية «القاموس العربي Arabic Dictionary»؛ وفكرته قائمة على التحليل الصرفي وتحويل النص المكتوب إلى صوت منطوق باستخدام تقنية التعرف على الصوت، بحيث إنك إذا أدخلت له نصًّا، وأردت أن تعرف معنى كلمة فيه، فإنه يعرض لك معناها وجذرها، كما ينطقها لك باستخدام تقنية تحويل النص إلى صوت، وهذا إلى حد كبير يفيدنا في مجال التعليم، لكنه ليس مبنيًّا على مُدَوّنات تراعِي الفروق الفرديّة والمعاييرِ التربويّة والتعليميّة؛ ليظلَّ محصورًا في إطار صلاحية كونه مرجعا وليس منهجا، فالمعاجم لا تُقَرَّرْ على الطلابِ، بل تكونُ مَرْجِعًا لهم، ونحنُ بحاجة إلى منهجٍ به نصوصٌ تُعَالَجُ حاسوبيا صوتيا وصرفيا ونحويا ودلاليا، ويكونُ الحاسوب له القدرةُ مثل الإنسانِ على تعليمِ مهاراتِ الكتابةِ والقراءةِ والتحدُّثِ والاستماع.

إن حاجتَنا إلى المعجم المعاصر لا تقتصرُ على رصدِ الكلماتِ الشائعةِ والـمُسْتَعْمَلَة، بل نحتاج إلى تطبيقات الذكاء الاصطناعيِّ للحصول على الكلمات من خلال تَقْنيَّة تحويل النص إلى كلام منطوقٍ، وبهذه الخَطْوة نلتقي في التعليمِ مع المرحلةِ السابقة التي أشرنا إليها في تقنية التعرُّفِ على الصوت، فيبدأ المتعلم بربطِ ما يسمعُه ويُكَرِّرُه بما يقرؤه من مفرداتٍ وألفاظٍ في هذا المعجم، خاصةً وأن في العربية إشكالات عدمِ التطابُقِ بين المكتوبِ والمنطوقِ، مثل (هذا) تُكْتَب بدونِ ألفٍ، لكنّها تنطق بألف، ومن

ثمَ فإننا نحتاجُ إلى معلوماتٍ صوتية عند كل مُدَخل مُعجمي للتسهيل على الدارسين، وتشملُ هذه المعلوماتُ: كيفية النطقِ مثل عَمْرو - عُمَرْ، وجنس الكلمة هل هي مذكر أم مؤنث، وعددها من حيثُ الإفرادُ والتثنيةُ والجمع، وما يتعلقُ بها من معلومات صرفية هل الكلمةُ تكونُ اسمًا أم فعلًا أم حرفًا وهل هي اسمُ فاعل أو مفعول.. إلخ، كما نحتاجُ في هذا المعجمِ التعليميِّ أن نُعَرِّفَ المتعلم هل هذه الكلمةُ عربية خالصة أم مُعَرَّبَة؛ وهكذا نُعْطِي المتعلم في هذا المعجم ذخيرة لغوية تتعلقُ بكلِّ ما يتصلُ بالكلمةِ بداية من نطقها وانتهاء بما يتعلق بها من معلوماتٍ وأوصافٍ لُغويَّة.

ولنا أن نتخيل أن هناك وجود منصة لـمُعجم مدرسي متاح لجميع الطلاب تساعدهم على قراءةِ الكلمةِ بشكلٍ صحيح وتُدرِّبهم على النطقِ، ثُمَّ تُعَرِّفُهُم بأحوالِ الكلمةِ الصرفيةِ وطريقةِ كتابتِها وتتيحُ لهم سياقاتِ استعمالِها من النصوص من خلال عرض النتائجِ الأكثرِ شُهْرَةً من المدوَّنة التي بُنِي منها هذا المعجم. إن هذه المنصةَ تجعلُ التطوير والتحديثَ في المناهجِ الدراسيةِ الـمُتَعلقةِ باللغةِ العربيةِ مُسْتمرًا وبشكلٍ مُنَظَّم يَبْني الملكاتِ النُّطقيةَ والقدراتِ العقليةَ، وأرى أنه بات فريضةً الوقت الذي نعيشُ فيه.

٧. الذكاء الاصطناعي والكتابة.

تعلم الكتابة يتطلب عدة مهاراتٍ، أولاها معرفة شكل الحرف ورسمه بالطريقة الصحيحة، والثانية ترجمة النطق السليم إلى كتابة صحيحة، والمستوى المهاري الثالث هو كيفية التأليف بين المفردات في جمل مفيدة. والذكاء الاصطناعي يمكن أن يقدم لنا حلولا في هذه المستويات الثلاث من خلال تقنية تقنية (OCR) التي تعني باختصار لكلمة (Optical character Recognition) وتعني التعرف الضوئي على الحروف، وهذه التقنية يمكن توظيفها في تعليم الخط والكتابة الصحيحة للحروف بنفس الفكرة المنطقية التي عرضت لتقنية التعرف على الصوت، فيمكننا بناء نموذج لخط النسخ بصفته أول خط يمكن تعلمه للأطفال، ثم بناء نموذج لخط الرقعة، ويطلب من الطلاب محاكاة هذا الخط على ورق عادي ثم مسح هذا الورق ضوئيا، فيُعرِّفه الحاسوب مواطن الخطإ وأبعاد الحروف وكيفية كتابتها على السطر وما إلى ذلك من إرشادات، ويعلمه طريقة كتابة الحرف منفصلا ثم متصلا في كلمات.

لقد حققت تقنية (التعرف الضوئي على الحروف) تقدما في معرفة الخطوط، لكن بها مشكلات في التعامل مع الخط العربي الذي يتميز بأنه يكتب بشكل متصل بخلاف الإنجليزية التي تكتب حروفها منفصلة، وهذا الاتصال يُصَّعِبُ على التقنية معرفة حروف الكلمة، لكن هذه المشكلة يمكن حلُّها بناء على اختيارِ أشهر الكلمات التي استخرجناها من مُدَوَّنَة الـمُعْجَم الذي أشرنا إليه في مرحلةِ بناءِ معجم معاصر، ثم تدريب التقنية على صورٍ مثاليةٍ ومعياريةٍ للحروف في هذه الكلمات، ثم جمع عينة من كتابات الطلاب واستخراج ملامح الخطإ في كتابة الحروف، وبشكل رياضيّ يمكن لهذه التقنية أن تتطورَ على الكلمات المختارة كبدايةٍ لمدوَّنَة صغيرةٍ لتعليم الخط تتسع شيئا فشيئا كلما زاد استعمال الطلاب لتقنية (التعرف الضوئي على الحروف).

إن اتصال الكلمات ليسَ المشكلة الوحيدة في تقنية (التعرف الضوئي على الحروف)، لكن هناك مشكلات أخرى تتمثل في نقط الحروف أو الإعجام، ونصف حروف العربية منقوط والنصف الآخر ليس منقوطا، وهذا ليس موجودا في الإنجليزية، ومن ثم قد تخطئ تقنية (التعرف الضوئي على الحروف) في التفرقة بين الحاء والخاء والجيم بسبب النقطة، والباء والياء والدال والذال، ومعالجة هذه المشكلة تتمثل في إعطاء التقنية الكم الأكبر من صور هذه الحروف ساعة اتصالها وانفصالها.

ومن المشكلات التي تواجه تقنية (التعرف الضوئي على الحروف) أيضا مشكلة التشكيل، فالإنجليزية ليست بها فتحة أو ضمة أو كسرة أو شدة أو سكون، أما العربية فإنها تعتمد بشكل أساسي على هذا التشكيل الذي نحتاجه لضبط النطق؛ والتقنية (التعرف الضوئي على الحروف) تخطئ في معرفة هذا التشكيل خاصة مع تعدد الخطوط من رقعة ونسخ وديواني وفارسي.. إلخ، والحل هو الاقتصار على نوع واحد من الخط في عدد محدود من الكلمات كما أسلفنا.

إن وضع قواعد محددة لكتابة الخط ليس أمرا صعبا، وهذه القواعد تتمثل في معرفة أشكال الحروف متصلة ومنفصلة وطريقة نقطها وتشكيلها، فمثلا ندرب الحاسوب على أن الحرف لا تُرسم عليه دائرة السكون إذا جاء في بداية الكلمة، ولا يمكن أن يجتمع حرفان ساكنان في كلمة، وغير ذلك من قواعد الكتابة والتشكيل.

إن هذه التقنية رغم مشكلاتها فإنها تمثل حجر أساس يمكن بناء صرح تعليمي ذكي عليها، لتعليم الخط الحسن الذي نفتقده في كثير من المتعلمين اليوم الذين يعتمدون على الكتابة الإلكترونية، ومعها يقل استعمالهم للقلم، رغم أن الكتابة اليدوية شيء مهم ولا يمكن الاستغناء عنه مهما تقدمت التكنولوجيا.

٨. الذكاء الاصطناعي وتعليم الإملاء.

تفيدُنا تَقنية (التعرف الضوئي على الحروف) أيضا في تعلم الإملاء حيث تُقارِنُ التقنية بين الكلماتِ المكتوبة في قاعدةِ البيانات وما يُنْتِجُه المتعلمُ من كلماتٍ تُمْلَى عليه، ويُحَدِّدُ له الْخَطَأَ فيها، هذا على مستوى الخطِّ اليَدَوي أو أيِّ خَطٍ آخَر أُدْخِلَ بشكلٍ ضَوْئِي.

إن التدقيقَّ الإملائي في الحاسوبِ خطا خطواتٍ متقدمةً من خلال تطبيقاتِ المدقق الإملائي، وأشهرها المدقق الإملائي لشركة مايكرو سوفت والمدقق الإملائي لشركة صخر. وفكرة المدّقِق الإملائيِّ تقوم على إيجادِ بدائلَ واحتمالاتٍ لتصحيحِ الكلمة الخطإ. ودقةُ هذين المدقِّقَين في الكلمات المفردة أعلى منها في الكلمات المركبة، ويتميز مدقق صخر عن مدقق (مايكرو سوفت) بإمكانية التصحيحِ التلقائي وتقليلِ عددِ الاقتراحات ومراعاةِ السياق والتعامل مع المصطلحات والتشكيل.

إن محركاتِ البحثِ المعنيَّة بالتدقيقِ الإملائيِّ الحاليةِ تُعاني من مشكلاتٍ كثيرة، فمثلا هناك تطبيقٌ يُسَمَّى: «اكتب صح ektebsa٧»، يُصَحِّحُ النصوصَ إملائيا، وبه قائمة من المتلازماتِ الأسلوبيةِ والأخطاءِ الشائعة، لكنْ عند تَجْرِبة هذا التطبيق بكتابة جُمْلة قصيرةٍ ظهرت به عدة مشكلات منها:

- أولا: لا يُفَرِّقُ المدقق في كثير من الأحيان بين الحروفِ الـمُعْجَمَة وغيرِ الـمُعْجَمَة، مثل الحاء والجيم، والدال والذال، والطاء والظاء.
- ثانيا: لا يستطيع اكتشاف الأخطاء النحوية في الكتابة إذا طالَ الفصلُ، فمثلا همزة (إنَّ) تُكسرُ في سبعةِ مواضع، لو جاء المثال: قالت: إنَّ أبي يدعوك، هنا لن يخطئ المدقق، لأنَّ النموذجَ الـمُدْخَل في قاعدة بياناتِه ينصُّ على هذا الشكل، أما لو جاء المثال: قالت ابنة شُعَيْبٍ لموسى: (إن) أبي يدعوك، هنا طالَ الفصلُ ولن يتمكنَ المدققُ الإملائي من التصحيحِ، فلو كتبنا (أن) ما اكتشفها.

إن التدقيق الإملائي يعتمدُ على فهم النص، وهذا يحتاجُ إلى خوارزميّة رياضية معقدة، ومُدَوَّنةٍ كبيرةٍ من خلالها نستطيعُ أن نضعَ القواعدَ التي من خلالها يمكننا الحصول على دقة عالية، لكن إنْ اقتصرَنا على تقنيةِ التعرُّف الضوئي من خلالِ رصدِ النماذج الصحيحة في الكتابة، ومقارنتِها بما يكتبُه الطلابُ، ثم بدأنا في توسعة المدونة بناء على ما نجدُه من أخطاء وما نضعُه من قواعد نحدِّث بها البرنامج، في هذه الحالة سيُبْنَى مُدقِّقٌ إملائيٌّ عالي الدَّقة.

ومن الأمانة أن نذكرَ أن من يتيحون هذه التطبيقاتِ لا يزعمونَ أن دقتَها عالية، بل يعترفون بأنها لا تزال قيدَ التطوير وأنها تقتصر على المشهور من الأخطاء فقط.

إننا بهذه التقنيات الثلاث يمكن أن نحققَ طفرة في تعليم العربية، فإذا افترضنا إمكانية تحويل الكلماتِ المكتوبةِ يدويًّا أو الممسوحةِ ضوئيا إلى نصٍّ رقمي، ثم أدخلنا هذه الكلمات على المدقق الإملائي والنحوي، إننا بذلك نحصلُ على نتيجةٍ تعليميَّةٍ هائلة، مع الاعتراف بأنَّ الدقةَ لن تكونَ عاليةً في البداية، ولكن كلما كثُرت المادةُ الـمُدْخَلَةُ للتقنية زادتْ دِقَّتُها.

إن تقنيةَ المدقق الإملائي لا تكتفي فقط بإعطاء الشكل الصحيحِ للكلمةِ إملائيا ونحويا، بل تعطيه معلوماتٍ عن هذا الخطإ الذي وَقَعَ فيه، فإذا كتب مثلا (سُئِلَ)، فإنَّ المدققَ يُعْطِيه احتمالات: سُئِلَ، أم سَأَل، ثم إذا وقف على الكلمة الأولى استخرج منها معلوماتٍ لغوية حول الكلمة منها: أن الهمزة تكتب على نبرة إذا كانت مكسورة، ومنها أن الفعل مبني للمجهول، وما يأتي بعده يكون نائبا للفاعل.. وهكذا في كل الكلمات.

وإذا افترضنا جدلا أننا أعطينا المتعلمَ نصًّا من خمسة أسطر، ودربناه بتقنية (التعرف على الصوت) على نطقه وأدائه، ثم أعطيناه معاني الكلمات من خلال المعجم الذي يُبنى تلقائيا باستعمال الدارسين، ثم طلبنا منه أن يقلّد الخطَ وبعدها قام الحاسوب بقراءة النصِّ أمَامه وأعطاه فرصة للاستماع والإملاء، ثم يعرضُ ما تعلَّمه على الحاسوب، فيقارنُ الخطوط ويضع له إرشاداتٍ لكتابةِ الكلمةِ بخطوطٍ صحيحة، ثم يُدَرِّبُه على الإملاء ويُعرِفُّه أخطاءه ويعطيه المعلوماتِ الوافيةَ حولَ الكلمة من حيثُ النطقُ والكتابةُ والمعنى المعجمي والاستعمالُ السياقي، إن تقنيةً بهذه الجودة لا بد وأنْ تُحْدِثَ

طَفْرَةً هائلةً في تعليمِ العربية، وهذه التقنيةُ ليست أضغاثَ أحلامٍ بل مشروعٌ قابلٌ للتطبيق، ولكنّه يحتاجُ فقط إلى التطوير المستمر، فأساسُ التقنية موجودٌ، ولكنَّ هذا الأساسَ يحتاج أن يُبْنى عليه.

٩. الذكاء الاصطناعي وتعليم النحو.

إذا وصلنا بالمتعلم إلى أنه صار قادرا على نطقِ الكلمات وكتابتِها ومعرفةِ معانيها، يُمكن توجيهه حينئذٍ إلى تعلم النحو لتركيب جمل مفيدة. والنحو لا يزالُ يمثل مشكلةً عند كثير من متعلمي العربية، والسبب في ذلك - من وجهةِ نظري- فصلُ القواعد النحويَّة عن النصوص، فصار النحو قواعدَ مُجردة تنسى سريعا، ولم يكن ملكة أو مهارة تُقَوِّمُ اللسان.

إن الذكاء الاصطناعي لم يصل في مرحلة تعليم النحو إلى تقدُّم يمكن وصفه بالكبير، مقارنة بالذكاء الاصطناعي في مجال الصوت والصرف والمعجم؛ لأن النحو يعتمد على الفهم، ويحتاج إلى فك اللبس الدلالي بين الكلمات لتحديد مواقعها الصحيحة في الجملة، فمثلا: جملة (قال الرجل) تتكون من فعل وفاعل، لا يمكن معرفة دلالتها إلا بإضافة مكمل، فنقول (قال الرجل كلمة الحق) فتكون (قال) من القول والكلام، أو نقول: (قال الرجل وقت القيلولة)، أي نام في وقت القيلولة والمعنى هنا تغيّر جذريا بسبب مكملات الجملة.

ورغم هذه المشكلات في النحو فإننا نعيد التأكيد على أن معالجة عينة أشهر الكلمات وبناء منصة تعليمية عليها يعد نواة لتلافي الأخطاء الصادرة عن التقنية وتكوين منصة يمكن تطويرها فيما بعد. فالمدقق النحوي سيعطي بناء على الكتابة الصحيحة والضبط السليم احتمالاتٍ أكثر دقة للمتعلم، فمثلا إذا كتب الدارس: (رأيت المتهمين) فالتطبيق يعطيه اقتراحات لضبط الميم: (رأيت المتهمِّين) أم (رأيت المتهمَّين)؛ ليفرق بين المثنى والجمع، أما إذا كتب (رأيت المتهمون)، فإن المدقق النحوي في كل الأحوال سيفيده أن (المتهمون) خطأ لأنها مفعول به منصوب وعلامة نصبه الياء لأنه جمع مذكر سالم أو مثنى حسب التحديد الذي سيكون أمامه في النص.

ويعتمد المدقق النحوي على السوابق واللواحق، فمثلا التركيب (لن يأكلون) لا يمكن أن تأتي بهذه الصورة؛ لأن (لن) من أدوات النصب، واللاحق (النون) لا يمكن أن تأتي مع لن الناصبة، وهنا تعطيه التقنية معلومة أدوات نصب الفعل المضارع المتمثلة في (أن - لن - كي - لام التعليل - حتى) كما تعطيه علامات النصب الأصلية والفرعية المتمثلة في الفتحة والألف والياء وحذف النون، وهكذا في كل الأدوات والحروف.

إن المدقق النحوي عندما يشرحُ كلَّ خطإٍ من خلال معلومات سريعة مباشرة، فإنها تكون أكثر رسوخا في ذهن المتعلم، وبكثرة التدريب ستستقر المعلومة في رأسه ويمكنه تلافيها بعد ذلك.

ولبناء تطبيق يعتمد على الذكاء الاصطناعي في تعليم النحو يمكننا أن نضع المنهج التالي:

١) اختيار أكثر دروس النحو استعمالا في الكلام.

٢) تصفية الدرس النحوي من الخلافات والكلمات الغريبة، والأساليب غير المستعملة، مثل ما انفك وما برح وغيرها.

٣) بناء نموذج لجمل سهلة لم تحدث فيها عوارض التركيب مثل الحذف والإضافة، والتقديم والتأخير.

٤) بناء نموذج للتوارد المعجمي والدلالي للتوافق في بناء الجملة، فعندما يكتب الطالب كلمة شرب فإن الحاسوب لا بد أن يحصر خياراته في أن يكون الفاعل كائنا حيا، فإذا كتب الطالب: شرب محمدٌ، فإن الحاسوب يحدد خيارات المفعول به في المفردات التي تصلح للشرب، فلا يأتي المفعول، الجبل أو الحائط، بل يكون شرب محمد العصير، أو شرب محمد الماء.. وهكذا

٥) تدريب الحاسوب على تحديد علامات الإعراب الأصلية والفرعية، وذلك اعتمادا على تقنية المحلل الصرفي، فمثلا تكون هناك معادلةٌ أركانُها: أن الفعل يأتي بعده الفاعل مرفوعا بالضمة إذا كان مفردا أو بالألف إذا كان مثنى أو بالواو إذا كان جمع مذكر سالم.

٦) تدريبُ الحاسوب على المتلازِمات الـمُقيَّدة والـمُركَّبات التي تأتي على صورة واحدة، وهي قليلة ويمكن حصرها، مثل: عدد من المركبات الاسمية كمُركَّب الصلة والموصول، ومُركَّب المصدر.

٧) محاولة التفرقة بين مركبات النعت والإضافة؛ لأنها أشهر المركبات استعمالا في الكلام.

إن عددا غيرَ قليل من قواعد النحو يمكنُ تحويلُها إلى معادلاتٍ رياضيةٍ من خلالِها نستطيعُ أن نقدِّمَ نموذجا معياريا صحيحا وندربَ الطلابَ على الجُمَل القصيرة كمرحلة أولى حتى نتقدم شيئا فشيئا من فكِّ اللبس الدلالي في الـمُدَقِّقَاتِ النحوية.

١٠. الذكاء الاصطناعي وإعادة بناء النصّ.

نَقْصِدُ بإعادة بناء النص هنا التلخيص، وإعادةِ ترتيب الكلام في الجملة بشكل صحيح. وتقنيةُ التلخيص تنقسم إلى نوعين: التلخيصِ الاستخراجيّ، ويتضمنُ اختيار عباراتٍ وجملٍ من المستند الأساسيّ؛ لتكوين الملخص الجديد دون إعادة صياغةِ هذه العبارات والجُمَل. والتقنياتُ المعتمِدَة على هذا النوعِ تتضمنُ تقييمَ أهميَّة الجُمَلِ والعبارات، ومن ثم اختيار تلك التي لها التقييم الأعلى على افتراض أنّها هي الجمل والعبارات التي تحوي المعنى الذي أراد الكاتب إيصاله.

أما التلخيصُ التوضيحي للنص فيتضمنُ تلخيصَ النصِّ بعبارات وجمل جديدة كليًّا؛ حيث يتم توليد جمل جديدةٍ تحوي الرسالة الأساسية للمستند الأصلي. ويعدُّ هذا الأسلوبُ أكثر تحديًا من الأسلوب السابق، ولكنه أكثر محاكاةً للأسلوب البشري.

إن تقنية التلخيص قام على تطويرها مجموعة من الباحثين منهم المهندس إبراهيم صبح في رسالته للماجستير (صبح، ٢٠٠٩). ويعتمد في هذه التقنية على التلخيص الاستخراجي، وفيه محاولة لحذف الجمل التي يراها غيرَ ضرورية للمتعلم، واعتمد في ذلك على عينة عشوائية من المقالات المنشورة في الصحف، وقد شَرُفتُ بمعاونتِه في هذا الجانب، حيث كنت أقُوم بحذفِ جُمل أراها غيرَ ضُروريّة في فهم الـمُحتوى، ثم تُدفَع هذه الـمُدْخلات اليدويّة إلى التقنية التي يُمكنُها بعد ذلك أن تستنبط ما يمكنُ حذفه في النص ولا يؤثرُ في معناه.

وأما إعادةُ إنتاجِ جُملٍ أو بنائِها فتعتمدُ على تقنياتِ التدقيقِ النحويِّ والتلخيصِ والتحليلِ الصرفي، لأننا سنعطي الكلماتِ سماتٍ حسْبَ مواقعِها في الجملة، فمثلا الكلماتُ النكرة لا يمكن أن نبدأ بها الجملة، والفاعلُ يكون مرفوعًا، والفعل الماضي يكونُ مبنيا على الفتح، والمصافُ إليه يكون مجرورًا وهكذا، فعند إعطاءِ المتعلمِ كلماتٍ ليطلب منه إنشاء جملةٍ صحيحةٍ مفيدةٍ فإن التقنية ستراعي كلَّ هذهِ المحددِّات سلفا. فمثلا الكلمات (المسلمون، رمضان، شهر، صام) نقوم بتقديمها للمتعلم ثم نطلب منه ترتيبَها وعند الخطإ نرشدُه إلى موقع الكلمة في الجملة.

إن الاستفادةَ من هذه التقنيةِ يمكنُ أن تساهمَ في تدريبِ الـمُتَعلِّم على رسمِ خارطةٍ ذهنية لما يقرؤه ويتعلَّمُه، فهو إذا تدرَّبَ على التلخيصِ الاستخراجيِّ فإنَّه يكتسبُ مجموعةً من المهاراتِ ستجعلُه قادرًا على الاستخراجِ التوضيحيِّ ومن ثَم بناء نصٍ جديد.

إنَّ آليةَ التلخيصِ الاستخراجيِّ تقومُ على تحويلِ النصِّ إلى شكلٍ رياضيٍّ تجري عليه مقاييسُ تَسْبِقُها قبلها معالجةُ للنصِّ، وتُحذفُ في هذه المعالجة الكلماتُ غيرُ المهمة، مثل الحروف والأدوات، وكذلك إعادة الكلمات إلى جذورها وقياس التشابه بين جمل النص وتحليل العلاقات الدلاليَّة بين الكلماتِ والجُمَل من خِلالِ مَصفوفَة رياضيّةٍ، وهناك طرقُ عدة لاستخراجِ الملخصِ من هذه المصفوفة، أبسطُ تلك الطرق، هو اختيارُ الجملِ ذاتِ الارتباطِ الأقوى بأهم المفاهيم. فمثلاً، لو أردنا أن يحتوي الملخصُ على عشر جُمَل، فإنا نختار الجُمَلَ العشر ذات القيم الأعلى في أول عشرة صفوف من المصفوفة.

إنَّ هذهِ التقنيةَ يُمْكِنُ استثمارُها في تطبيقاتِ الفهمِ والاستيعابِ للنص، بحيثُ يمكن أن نذكرَ للمتعلمِ بعضَ المعلوماتِ وهو يُكْمِلُها، وهناك برامِجُ منتشرةٌ في عددٍ من الدولِ تُسَمَّى برامج كشفِ النصِّ المخفي، وتستخدمُ لتدريبِ الطالبِ على التنبؤِ بالنصِّ الُّلغوي الـمَخْفِي، وهذه النصوصُ تكونُ دُرُوسًا في القراءةِ والاستيعابِ أو قِطَعًا أدبيّةً من النثرِ والشعرِ، ويُصَمَّمُ البرنامجُ على شكلِ لُعْبَةٍ تَرْبَوِيَّة تحتسبُ فيها نقاطٌ عند التنبؤ بالمحذوفِ في الوقت الـمُحدد.

١١ . مشكلات الذكاء الاصطناعي في التعامل مع اللغة العربية.

تمثل مُعالجةُ اللغة العربية تحديا كبيرا، وهذا التحدي تختلفُ مستوياتُه، ففي الجانب الصوتيّ التحدي أقل؛ لأنَّ المعلومات الصوتية في اللغة العربية أكثر تحديدًا من المعلومات الصرفيّة ثم المعجميّة ثم النحويَة والدلاليّة، وفيما يأتي تحديد لبعض هذه المشكلات.

١) اللبس الدلالي.

العقبة الرئيسيّةُ التي تصادِفُنَا هي اللبس الدلاليّ سواء على مستوى البنية أم التركيب، فاللبسُ يقفُ عائقًا دونَ تحقيقِ دقةٍ عاليةٍ في معالجةِ النصوصِ العربيةِ، فمثلا كلمةُ الولَدان (مثنى) كلمة (ولد) تلتبس بالوِلْدان وهي جمع، والأولى تُرفعُ بالألف وتُنصبُ بالياء، وأما الثانية فتكونُ على حالِها رفعًا ونصبًا وجرًّا، ولا يستطيعُ الحاسوبُ أن يفرقَ بين (بالولَدَان) و (بالوِلْدان)، بل يَخلط الحاسوبُ بين فطائر جمع فطيرة، وبين فطائر مفرد طائر التي ألصقت بها الفاء كسابق (جمعة، ١٤٣٧ : ٧٤).

وعلينا أن نعترفَ أن الدراسةَ البحثيَّةَ في معالجةِ الدلالة لا تزالُ تحتاجُ الكثيرَ من الجهودِ الـمُكَلِّفَةِ ماديًّا، فلا يُمكنُ للقطاعِ الخاصِّ أن يتحملَها، ومن ثَمَّ لا بد من مشروعٍ كبيرٍ يضمُّ صَفْوَةَ الباحثين لتطويرِ التقنياتِ التي توصلنَا إليها وتلافِي الأخطاء التي تقعُ فيها هذه التقنية.

٢) طريقة بناء الجملة العربية.

طريقةُ بناء الجملة العربيةِ تحتاجُ إلى عمليةٍ مُعَقَّدَةٍ من الفهم ومَعْرِفَةِ سِمَاتِ المفرداتِ التي تناسبُ الوظيفةَ النحويَّةَ، بالإضافةِ إلى أنَّ الجُمْلَةَ العربيَّةَ تَحدُثُ لها عَوارِضُ تركيبٍ من الحذفِ والإضافةِ والتقديمِ والتأخير. وهذا يُمَثِّلُ تحديًّا كبيرًا للذكاءِ الاصطناعي.

٣) تنوع اللهجات.

تنوعُ اللهجاتِ العربيَّةِ يتحدَّى تقنياتِ التعرُّفِ على الصوتِ، خاصة فيما يتعلق بإبدال الحروف (السين مكان الثاء والجيم مكان القاف أو الغين مكان القاف، أو الظاء مكان الضاد)، هذا فضلا عن اختلافِ النبر والتنغيم وأداء الكلام، وهذا التنوعُ اللهجيُّ ليسَ موجودًا في اللغاتِ الأخرى بهذه الكثرة والتنوع.

١٢. مقترح لمنصة تعليمية ذكية.

بعد هذا العرض فإنني أقترحُ مِنَصَّة تعليميَّة للغة العربية تجمع بين كل هذه التقنيات، لتكون نواة لمدونة تعليمة كبرى تحتوي على أصوات اللغة ومفرداتها وتراكيبها، وهذه المنصة يمكن بناؤها من خلال صناعة مدونة لأشهر المستعمَل في اللغة العربية.

ويمكن تطبيقها أيضا من خلال تدريب كل التقنياتِ الحاسوبية على هذه المدونة، بمعنى أننا سنحاول إنشاء نصوص موجهة باستخدام مفردات هذه الـمُدَوَّنَة ومسكوكاتها فقط؛ حتى نتجنب المشكلات التي تقع فيها التقنية، وحتى نضمن دقة عالية في المعلومة التي تصل إلى المتعلم.

وكإطار شكلي يمكن تصور هذه المنصة على النحو التالي:

كما في (الشَّكل ٣) نختارُ جملةً من النصوص ثم نبدأُ في معالجَتِها لُغويًّا بشكل كامل، حيث نبدأُ بالتهيئةِ الحافزةِ التي قد تكون قصَّة أو مقطعًا مرئيا أو فكرة مُبْتَكَرة عند المعلم، والمنصة تتيح بعض هذه التهيئاتِ للاستعانةِ بها، وتترك المجالَ للمعلم كي يضيف فكرتَه الإبداعيةَ في هذه التهيئةِ التي ستضمُّ تَجارِبَ المعلمين وخبراتِهم، فيحدُثُ احتكاكٌ وتتوالد الخبرات التي تفيد المتعلمين، ثم تبدأ المعالجةُ اللغوية للنص، فيقرأ النصَّ بصوت جميل، ويستمع الطلابُ لأداء نوعيٍّ للنص، ثم يُسمَح لهم بقراءةٍ جهريَّةٍ على إِثْرِها تُسْتَخْدَمُ تقنية (التَّعَرُّفِ على الصَّوْتِ) حيث يستمع الطالبُ لصوته، وتبدأ التقنية في تظليل الكلمات التي أخطأ في نطقِها صوتيا أو صرفيا أو نحويا، ثم بالإشارَة إلى الكلمةِ أو الجملة تظهرُ الرسائل التوضيحية والتعليمية.

وبعد هذا التدريبِ على القراءة يكون الطالبُ قد استوعب فكرة النص، ويكون جاهزا للإجابة عن أسئلة الاستيعاب التي تأتي في هذه النافذة؛ حيث يمكن استثمارُ تقنية التلخيص في أن يُطْلَبَ تلخيصُ النص الذي قرأه أو التعبيرُ عنه بأسلوبه.

كما تتيح هذه النافذة أن يمسح النص، ثم تبدأ التقنية بقراءته بشكل بطيء يسمح للطالب أن يكتبه إملاءً، وبعد أن ينتهي يقوم بإدخال خطِّه مُصوَّرًا، ومن خلال تقنيةِ التعرف الضوئي يُمْكِنُ أن يكتشفَ الحاسوبُ مواطنَ الخطإ ويُعْرَضُ له النصُّ الأصليُّ مع النص الـمُمْلَى.

وتتيح النافذة أيضا أسئلة تقييم تُسْتَخْدَمُ فيها الوسائلُ التربويةُ في وضعِ الأسئلةِ من تحليل وتركيب وتقعيد.



الشكل ٣: نموذج المنصَّة التعليميَّة - نافذة مُعطيات رئيسية

تكون بالمنصة مدونة مصنفة ومرتبة حسب شيوع كلماتها، وبناء على المدونة، نختار مثلا الآيات القرآنية ذات الكلمات المتداولة والأحاديث ذات الجمل البسيطة والشعر الذي يستخدم كلمات مألوفة للأذن وهكذا في كل النصوص. وهناك إمكانية في المنصة ليضيف المعلم والمتعلم ما يروق لهما من نصوص، وهذه المختارات يمكن أن ندرس طبيعتها وتوجهاتها وتكون قيمة مضافة للمنصة.



الشكل ٤: نموذج المنصَّة التعليميَّة - نافذة اختيار النَّصّ

وفي هذه المنصة توجد نافذة لتعلُّم المهارات التي نوظِّفُ فيها التقنيات التي ذكرناها في هذا الفصل، فمثلا إذا دخل على نافذة الإملاء يَظْهرُ له نموذج ليحاكيه، ثم يكتب بالقلم العادي على ورقة تُصَوَّرُ بشكل آلي وتُقَارَنُ خطُوطُها بخطوطِ النموذج، فيكتشفُ الـمُعلمُ أخطاءه الإملائية، كما يتعرفُ أيضا على طريقة رسمِ الحرفِ بشكلِه المعياريِّ سواء في الرُّقْعَة أو النسخ باعتبارهما من أشهر الخطوطِ الـمُسْتَعْمَلَةِ في العربية.



الشكل ٥: نموذج المنصَّة التعليميَّة – نافذة المهارات

وتشتمل المنصة أيضًا على مجموعة من معارفَ لُغويَّة تضمُّ خلاصاتٍ في الأصوات والصرف والنحو والدلالة، كما تضمُّ معارفَ ثقافيةً، ويمكن أن تُمَثِّلَ مُدَوْنَةً يستطيعُ المستخدمُ أن يُضيفَ إليها، مع وجودِ آليةٍ لدراسة مُخْتاراتِ المتعلمين واعتمادِها ضمن المنصة.



الشكل ٦: نموذج المنصَّة التعليميَّة – نافذة المعارف

وتشمل هذه المنصةُ أيضا وجدانياتٍ ترتقي بمشاعر المتعلم، وتُبْنى مدونتُها من خلال مختاراتٍ تناسبُ المستويات والأعمار والثقافات، ولا تقتصرُ على الثقافةِ العربية، بل تمتدُّ لثقافات أخرى، ويتاحُ للمتعلمِ أن يُضيفَ عليها، أو أن يُبْدِي رأيه من خلال قاعدةِ بياناتٍ يمكن استثمارُها في دراسَةٍ بحثيَّةٍ تبيِّنُ معرفةِ طريقة التفكير وأسس الاختيار التي ستختلفُ حَسْبَ البيئات والثقافات التي تتعدد في وطننا العربي، ويمكن استثمار هذه النافذة في معرفة العلاقات بين الألفاظ في ذهن المتلقي، بحيث نُنْشِئُ شبكة من الكلمات التي تنتمي لحقل دلالي واحد، وما بينها من علاقاتٍ دلالية كالترادف والتضاد والمشترك اللفظي وغير ذلك من العلاقات.

إن الطلاب والمعلمين يُمْكِنُهُم أن يُنْشِئوا قاعدة بياناتٍ تفيد الباحثين، وهذه القاعدةُ تَنْشَأُ بشكلٍ طبيعيٍّ وفي بيئاتٍ مُخْتَلِفَةٍ، فتمثل مادةً خَصْبَةً يُمْكِنُ دِرَاسُتها والبناءُ عليها.



الشكل ٧: نموذج المنصَّة التعليميَّة – نافذة وجدانيات

وتشمل أيضًا هذه المنصةُ وسائلَ التقييم حسب مستوى المتعلم، بدايةً من المرحلةِ الابتدائيةِ، ولكن في صورةِ مستويات. فمن الممكن أن يكون الطالبُ في الصفِّ الخامسِ، لكنَّه يظلُّ في المستوى الأول الذي قد يشتركُ فيه مع طالبٍ في الصف الثالث، وهذه المنصة سَتُحَسِّنُ من مستوى الطالب، وسترفعُ عنه الخجلَ من مرافقةِ طالبٍ أصغر منه سنًّا.

والمنصة أيضا بها قصص تثري معارف الطالب، كما أن بها مقاطعَ مرئية ومسموعة لتنويع أدوات التعلم، وكل هذا في مكان واحد وبأسلوب سهل الاستخدام.



الشكل ٨: نموذج المنصَّة التعليميَّة - نافذة التقييم

إن هذه المنصة تُعدُّ واجهة يمكن من خلالِها توظيفُ التقنيات التي أشرنا إليها؛ لتكون بيئةً تعليميَّة مثاليةً يمكن استخدامها في المدرسة أو البيت أو أي مكان.

الخلاصة.

خلاصة القول في هذه القضية أن برامج معالجة اللغات الطبيعية حقَّقَتْ طفرةً لا يمكن إنكارُها، وهذه الطفرة يمكن استثمارُها في تقنياتِ تعليم اللغة العربية، ولكنَّ الأمرَ يحتاجُ إلى مشروعٍ كبير يضمُّ صفوة الباحثين في الوطن العربي، كما أن هذا المشروع الذي يوفر منصة تعليمية متصلة بالمدارس في مختلف أنحاء الوطن العربي يمكن أن يُستثمر في تعليم الطلاب وفي تكوين قاعدة بيانات ضخمة، ويمكن أن يُكوِّن تغذياتٍ راجعةً تفيده في تطوير التقنيات التي يوظفها، حيث يتيح نافذة لإبداء الرأي في التقنيات والمشكلات التي تواجه المتعلمين، وبهذه الطريقة يمكن أن تكون هذه المنصة منهجا تعليميا مستقلا يوفر جهود طباعة الكتب ويعطي نتائجَ نوعية في تعليم الطلاب اللغة العربية.

ببليوجرافيا مرجعيَّة.

١. ابن جني، أبو الفتح عثمان، (٢٠٠٦) الخصائص، دار الكتب المصرية، تحقيق، محمد علي النجار.

٢. جرين، جوديث، (١٩٩٢)، التفكير واللغة، الهيئة المصرية العامة للكتاب.

٣. جمعة، عمرو (٢٠١٦)، تقنيات اللغة العربية الحاسوبية، معايير التقييم ورؤى التطوير، دراسة لغوية حاسوبية، مركز الملك عبد الله لخدمة اللغة العربية.

٤. حامد، صلاح الدين، (٢٠٠٤)، خوارزمية التعرف على أخطاء التلاوة، دكتوراه، ٢٠٠٤، كلية الهندسة جامعة القاهرة.

٥. دي سوسير، فردينان دي سوسير (١٩٨٥)، دروس في الألسنية العامة، تعريب محمد شاوش وآخرون، الدار العربية للكتاب، تونس، د.ط.

٦. السعيد، المعتز بالله (٢٠٠٧) مدونة معجم عربي معاصر، معالجة لغوية حاسوبية، ماجستير، جامعة القاهرة.

٧. السعيد، المعتز بالله (٢٠٠٩)، المعجم التكراري لألفاظ القرآن الكريم، «المنهج والنموذج» ندوة القرآن الكريم وتقنية المعلومات، المدينة المنورة، السعودية.

٨. السعيد، المعتز بالله (٢٠١٦)، توظيف الـمُدَوَّنات اللُّغَوِيَّة في تطوير مُقرَّرات اللُّغة العَرَبيَّة لمراحل التَّعليم العامّ، مجلَّة التَّخطيط والسِّياسة اللُّغويَّة، مركز الملك عبد الله لخدمة اللُّغة العربيَّة، الرِّياض، العدد ٣، ٢٠١٦.

٩. صبح، إبراهيم (٢٠٠٩)، التلخيص الآلي، ماجستير، كلية الهندسة بجامعة القاهرة.

١٠. طعيمة، رشدي (٢٠٠٦)، معاهد تعليم اللغة العربية لغير الناطقين بها، اتجاهات التطوير، معايير الاعتماد، مؤشرات الجودة، مجلة العربية للناطقين بغيرها، جامعة أفريقيا العالمية، العدد الثالث.

١١. عبد الغني، نعيم محمد (٢٠١٥) تفاعل المفردات مع الوظائف النحوية في الجملة القرآنية، دار النابغة، مصر، ط: ١.

١٢. علي، نبيل (١٩٨٧)، اللغة العربية والحاسوب، مجلة عالم الفكر، المجلد الثامن عشر، العدد الثالث.

١٣. علي، نبيل (١٩٨٨)، اللغة العربية والحاسوب، تعريب، القاهرة.

١٤. عمايرة، خليل عمايرة (١٩٨٥)، في التحليل اللغوي، مكتبة المنارة، ط: ١.

١٥. عمر، أحمد مختار (١٩٨٥)، علم الدلالة، عالم الكتب، القاهرة، ط:٢.

1. Al-sulaiti, L. (2004). *Designing and Developing a corpus of contemporary Arabic*. Leeds University.
2. Chomsky, N. (2000). *New Horizon in the Study of language and mind*. Cambridge University Press.

الباحثون



الدكتور/ مُحمَّد عَطِيَّة مُحمَّد العَرَبيّ

عَمِلَ «د. محمد عطية محمد العربي أحمد» لمدة ربع قرن في مجال علوم الحاسب، وكذلك علومِ اللسانيات الحاسوبية، وحوسبةِ اللغة العربية، وبناءِ الموارد اللغوية. وقد نشر العديدَ من الأوراق البحثية باللغة الإنجليزية في الصف الأول من الدوريات العالمية المحكَّمة، كما حَكَّمَ العشرات من الأوراق في تلك الدوريات، وشارك في عدد من الكتب والدراسات والأوراق البحثية بالعربية حول المعجم، والصرف، والتركيب، والدلالة في اللغة العربية. وكذلك شارك وأدار العديد من التقنيات الحاسوبية الصناعية، والمشروعات البحثية في حوسبة اللغة العربية، والتعليم الإلكتروني، وكان خبيرًا حاسوبيًّا لمشروع المعْجَم التاريخي للغة العربية بين عامي ٢٠١٤م و ٢٠١٦م. وكذلك كان أستاذً زائرًا في عدة أكاديمياتٍ عربيةٍ مثل: الأكاديمية البحرية للعلوم والتكنولوجيا والنقل البحري، وجامعة الإسكندرية.



الدكتور/ المُعتزّ بالله السَّعيد طه

أستاذ الدِّراسات اللُّغويَّة الـمُساعد بجامعة القاهرة، وأستاذ اللِّسانيَّات الحاسوبيَّة المُشارك بمعهد الدَّوحة للدِّراسات العُليا، ومُنسِّق وَحدة الموارد الـمُعجميَّة بمشروع مُعجم الدَّوحة؛ وهو استشاريٌّ لعددٍ من الهيئات والمؤسَّسات. نَشَرَ نحوَ ثلاثينَ ورقة علميَّة مُحكَّمة، بالإضافة إلى عددٍ من الكتب في حوسبة اللُّغة وصناعة الـمُعجم والدَّرس اللُّغويّ الـمُعاصِر، وقامَ بتحكيم عشرات الأبحاث والأطروحات العلميَّة. شاركَ في العديد من المشروعات الدَّوليَّة المعنيَّة بمُعالَجَة اللُّغات الطَّبيعيَّة، وأشرفَ على بناء وتطوير الموارد اللُّغويَّة والحاسوبيَّة لعددٍ من هذه المشروعات. حصلَ على عددٍ من الجوائز العلميَّة، منها: جائزة (ألكسو ALECSO) للإبداع والابتكار في «المَعلُوماتيَّة والـمُعالَجَة الآليَّة للُّغة العربيَّة»، وجائزة راشد بن حُمَيد للعُلُوم والثَّقافة.



الدكتور/ أحمد راغب أحمد

أُسْتَاذ الدِّراسات اللُّغَوِيَّة المُشَارِك، ورَئِيس قِسْم اللغة العربية وآدابها بالجامعة الإسلامية العالمية بماليزيا. حصلَ على درجة الدُّكتوراه بمرتبة الشَّرف الأولى من جامعة القاهِرة عام ٢٠٠٩، وَنَشَرَ عددًا من الأوراق البحثِيَّة في دوريَّاتٍ علميَّة؛ كما شارَكَ في العديد من المؤتمرات الدَّوليَّة المعنِيَّة بحوسبة اللُّغة؛ وله عددٌ من المؤَلَّفات العِلميَّة.



الدكتور/ نعيم مُحمَّد عبد الغني

حصلَ على الدُّكتوراه في النَّحو والصَّرف والعَرُوض بمرتبة الشَّرف الأولى من جامعة القاهرة. يعملُ في البحث العلميّ والتدريسِ الجامعيّ والإعلام، وهو الآنَ خبيرٌ لغويٌّ في المنظَّمة الإسلاميَّة للتَّربية والعُلُوم والثَّقافة (إيسسكو)؛ وقد عملَ مُحاضرًا في جامعة قطر، وباحثًا غيرَ مُتفرِّغ في مُعجم الدَّوحة التَّاريخيّ، وكاتبًا ومُحرِّرًا في بعض المجلَّات والإذاعات العربيَّة. نُشِرَت لهُ عَشراتُ المقالات والتَّقارير في صُحُفٍ ومجلَّاتٍ مُختلفة، وأعدَّ وقدَّمَ العديدَ من البرامج الإذاعيَّة الَّتي تُعنى بالشَّأن الثَّقافيّ، وله ستَّة كُتُب مطبوعة، بالإضافة إلى عددٍ من الأبحاث العلميَّة المُحكَّمة.

مباحث لغوية ٦٠

العربيَّة والذَّكاء الاصطناعيّ

يُصدِر مركز الملك عبدالله بن عبدالعزيز الدولي لخدمة اللغة العربية هذا الكتاب ضمن سلسلة (مباحث لغوية)، وذلك وفق خطة عمل مقسمة إلى مراحل، لموضوعات علمية رأى المركز حاجة المكتبة اللغوية العربية إليها، أو إلى بدء النشاط البحثي فيها، واجتهد في استكتاب نخبة من المحررين والمؤلفين للنهوض بعنوانات هذه السلسلة على أكمل وجه.

ويهدف المركز من وراء ذلك إلى تنشيط العمل في المجالات التي تُنَبّه إليها هذه السلسلة، سواء أكان العمل علميا بحثيا، أم عمليا تنفيذيا، ويدعو المركز الباحثين كافة من أنحاء العالم إلى المساهمة في هذه السلسلة.

وتودّ الأمانة العامة أن تشيد بجهد السادة المؤلفين، وجهد محرر الكتاب، على ما تفضلوا به من رؤى وأفكار لخدمة العربية في هذا السياق البحثي.

والشكر والتقدير الوافر لمعالي وزير التعليم المشرف العام على المركز، الذي يحث على كل ما من شأنه تثبيت الهوية اللغوية العربية، وتمتينها، وفق رؤية استشرافية محققة لتوجيهات قيادتنا الحكيمة.

والدعوة موجّهة إلى جميع المختصين والمهتمين للتواصل مع المركز؛ لبناء المشروعات العلمية، وتكثيف الجهود، والتكامل نحو تمكين لغتنا العربية، وتحقيق وجودها السامي في مجالات الحياة.

الأمين العام للمركز

أ.د. محمود إسماعيل صالح



ص.ب ١٢٥٠٠ الرياض ١١٤٧٣
هاتف:٠٠٩٦٦١١٢٥٨٧٢٦٨ - ٠٠٩٦٦١١٢٥٨١٠٨٢
البريد الإليكتروني: nashr@kaica.org.sa



